A3: Rings and Modules, 2021–2022

Tom Sanders

We begin with the course overview as described on `https://courses.maths.ox.ac.uk/course/view.php?id=75`.

Course Overview:

The first abstract algebraic objects which are normally studied are groups, which arise naturally from the study of symmetries. The focus of this course is on rings, which generalise the kind of algebraic structure possessed by the integers: a ring has two operations, addition and multiplication, which interact in the usual way. The course begins by studying the fundamental concepts of rings (already met briefly in core Algebra): what are maps between them, when are two rings isomorphic *etc.* much as was done for groups. As an application, we get a general procedure for building fields, generalising the way one constructs the complex numbers from the reals. We then begin to study the question of factorization in rings, and find a class of rings, known as Unique Factorization Domains, where any element can be written uniquely as a product of prime elements generalising the case of the integers. Finally, we study modules, which roughly means we study linear algebra over certain rings rather than fields. This turns out to have powerful applications to ordinary linear algebra and to abelian groups.

Learning Outcomes:

Students should become familiar with rings and fields, and understand the structure theory of modules over a Euclidean domain along with its implications. The material underpins many later courses in algebra and number theory, and thus should give students a good background for studying these more advanced topics.

Course Synopsis:

Recap on rings[1] (not necessarily commutative[2]) and examples: $\mathbb{Z}$[3], fields[4], polynomial rings[5] (in more than one variable[6]), matrix rings[7]. Zero-divisors[8], integral domains[9]. Units[10]. The characteristic of a ring[11]. Discussion of fields of fractions and their characterization[12] (proofs non-examinable). [2]

Homomorphisms of rings[13]. Quotient rings[14], ideals[15] and the first isomorphism theorem[16] and consequences[17][18], *e.g.* Chinese remainder theorem[19]. Relation[20] between ideals in R and R/I. Prime[21] ideals and maximal[22] ideals, relation to fields[23] and integral domains[24]. Examples of ideals[25][26][27]. Application of quotients to constructing fields by adjunction of elements[28]; examples to include $\mathbb{C} = \mathbb{R}[X]/\langle X^2+1\rangle$[29] and some finite fields[30][31]. Degree of a field extension[32], the tower law[33]. [4]

Euclidean domains[34]. Examples[35][36][37]. Principal Ideal Domains[38]. EDs are PIDs[39]. Unique factorisation for PIDs[40]. Gauss's Lemma[41] and Eisenstein's Criterion for irreducibility[42]. [3]

Modules[43]: Definition and examples: vector spaces[44], abelian[†] groups[45], vector spaces with an endomorphism[46]. Submodules[47] and quotient modules[48] and direct sums[49]. The first isomorphism theorem[50]. [2]

Row and column operations on matrices over a ring[51]. Equivalence of matrices[52]. Smith Normal form of matrices over a Euclidean domain[53]. [1.5]

Free[54] modules and presentations of finitely generated modules[55]. Structure of finitely generated modules over a Euclidean domain[56][57]. [2]

Application to rational canonical form[58] and Jordan normal form[59] for matrices, and structure of finitely generated abelian [†] groups[60]. [1.5]

The links to sections of the notes above are intended as a starting point for the topic and are not exhaustive.

[†]We use the word commutative instead of abelian in these notes.

1 Rings: a recap

A set R containing two (possibly equal) elements 0 and 1, and supporting two binary operations + and × is a **ring** if

- R equipped with + is a commutative group with identity 0;
- × is an associative binary operation on R with identity 1;
- × is distributive over +.

Occasionally we shall have multiple rings and it will be instructive to clarify which particular ring we are referring to. We shall do this with subscripts writing, for example, $+_R$ or 1_R instead of + and 1 above.

The operation × is the **multiplication** of the ring, and we write xy in place of $x \times y$. We say R is a **commutative** ring if the multiplication is commutative. ⚠ The modern notion of commutative ring can be traced back to Emmy Noether [Noe21, §1] (translated into English in [Ber14]), though her definition does *not* assume the multiplication has an identity.

Remark 1.1. The multiplication is the most changeable part of a ring. We do *not* assume our rings are commutative, while we *do* assume they have a multiplicative identity. Poonen [Poo19] provides a full-throated defence of this last position.

There are also interesting non-associative rings both with identity *e.g.* the octonions, and without *e.g.* Lie algebras, but these are outside the scope of this course.

Remark 1.2. Since multiplication is an associative binary operation it can be shown (see *e.g.* [Hun80, Theorem 1.6, p28]) that for $x_1, \dots, x_n \in R$, the expression $x_1 \cdots x_n$ gives the same result regardless of how we insert brackets, provided we do so in a grammatically valid way; we take the unbracketed expression to denote this common result. The empty product, that is the product with no terms in it, is defined to be the multiplicative identity.

For a bijection $\sigma : \{1, \dots, n\} \to I$ and elements $x_i \in R$ for each $i \in I$, we write $\prod_{i=\sigma(1)}^{\sigma(n)} x_i$ for the product $x_{\sigma(1)} \cdots x_{\sigma(n)}$. If multiplication is commutative then it can further be shown (see *e.g.* [Hun80, Corollary 1.7, p28], though in this case the proof is left as an exercise) that $x_{\sigma(1)} \cdots x_{\sigma(n)}$ gives the same result for any bijection $\sigma : \{1, \dots, n\} \to I$; we write $\prod_{i \in I} x_i$ for this common result.

The operation + is the **addition** of the ring, 0 is the **zero** of the ring, and the set R with the operation + (and identity 0) is the **additive group** of the ring. For each $x \in R$ we write $-x$ for the unique inverse of x, and the map $R \to R; x \mapsto -x$ is the **negation** of the ring.

Remark 1.3. Identities are self-inverse so $-0 = 0$, and double inversion is the identity map, so $-(-x) = x$ for all $x \in R$.

Remark 1.4. A group operation is commutative if and only if inversion is a homomorphism of the group. In our particular case addition is commutative so negation is a homomorphism.

Remark 1.5. Since addition is an associative and commutative binary operation with identity, for a set I of size n, and elements $x_i \in R$ for each $i \in I$, the result of $x_{\sigma(1)} + \cdots + x_{\sigma(n)}$ for $\sigma : \{1, \ldots, n\} \to I$ a bijection is independent of how we put in brackets and the particular choice of bijection σ; we denote it $\sum_{i \in I} x_i$ and sometimes $\sum_{i=\sigma(1)}^{\sigma(n)} x_i$. The empty sum, that is a sum with no terms or, equivalently, a sum indexed over the empty set, is defined to be the zero of the ring.

Remark 1.6. Suppose I is a finite set and $x_i \in R$ for each $i \in I$. If $J \subset I$ is such that $x_i = 0$ for all $i \in I \setminus J$ then $\sum_{j \in J} x_j = \sum_{i \in I} x_i$; in words 'sums of zeros are zero'. In particular if $x_i = 0$ for all $i \in I$ then $\sum_{i \in I} x_i = 0$.

Remark 1.7. Suppose I is a finite set and $x_i \in R$ for each $i \in I$. Addition is commutative and associative so

$$\sum_{i \in I} x_i = \sum_{P \in \mathcal{P}} \left(\sum_{i \in P} x_i \right) \text{ for all partitions } \mathcal{P} \text{ of } I.$$

The fact that the right hand side is the same whatever the partition lets us swap between different partitions in a process called 'change of variables'.

There are many examples of the power of change of variables: given an $m \times n$ matrix of elements of R we can sum them up first by summing the rows, and then summing the row totals; or first by summing the columns, and then summing the column totals. The fact that these are the same already gives that multiplication of natural numbers is commutative, but there are many more applications, for example the proof of Burnside's Lemma in [Ear14, Theorem 286]; or the proof of the formula $1 + \cdots + n = \frac{1}{2}n(n+1)$ (which also uses the fact we can reverse the order of summation); or the proof of the Handshaking Lemma in [Lac20, Lemma 10], famous for its application to the Königsberg Bridge Problem. For us it plays an essential role in the proof of Proposition 1.12 (though we omit this proof!) and Proposition 1.60.

Given $y \in R$, the map $R \to R; x \mapsto yx$ (resp. $R \to R; x \mapsto xy$) is called **left** (resp. **right**) **multiplication by** y. The fact that multiplication is distributive over addition in R is exactly to say that all the left and right multiplication maps are group homomorphisms of the additive group of R.

Remark 1.8. Group homomorphisms preserve identities, meaning the identity of the domain is mapped to the identity of the codomain. In our case this means $x0 = 0x = 0$ for all $x \in R$ – we say 'zero annihilates'. Group homomorphisms also preserve inverses, meaning that the inverse of the image of x is the image of the inverse of x. Again, for our case we have $x(-y) = (-x)y = -(xy)$ for all $x, y \in R$.

Remark 1.9. The fact that all left *and* right multiplication maps are homomorphisms and multiplication has an identity (and the additive group is a group) entails the commutativity of the additive group. Exercise I.1 asks for a proof of this.

Remark 1.10. We follow the conventions of arithmetic in the integers by writing $x - y$ for $x + (-y)$, and in the absence of brackets ring multiplication precedes ring addition, for example $xy + z$ means $(xy) + z$.

Remark 1.11. For $A \subset R$ we write $-A := \{-a : a \in A\}$. If $x \in R$ then we write $A + x = \{a + x : a \in A\}$; $xA = \{xa : a \in A\}$; and $Ax := \{ax : a \in A\}$. Finally, if $B \subset R$ we write $A + B := \{a + b : a \in A, b \in B\}$ so in particular $A + x = A + \{x\}$.

Proposition 1.12 (Algebra of polynomials)**.** *Suppose that R is a ring, $a_0, a_1, \ldots, b_0, b_1, \cdots \in R$ have $a_i = 0$ for all $i > n$ and $b_j = 0$ for all $j > m$, and $\lambda \in R$. Then*

$$\left(\sum_{i=0}^{n} a_i \lambda^i\right) + \left(\sum_{j=0}^{m} b_j \lambda^j\right) = \sum_{i=0}^{\max\{n,m\}} (a_i + b_i)\lambda^i \text{ and } -\left(\sum_{i=0}^{n} a_i \lambda^i\right) = \sum_{i=0}^{n} (-a_i)\lambda^i.$$

Furthermore, if $\lambda b_j = b_j \lambda$ for all j, then

$$\left(\sum_{i=0}^{n} a_i \lambda^i\right)\left(\sum_{j=0}^{m} b_j \lambda^j\right) = \sum_{k=0}^{n+m} \left(\sum_{j=0}^{k} a_{k-j} b_j\right) \lambda^k.$$

Remark 1.13. We omit the proof though it is not difficult: it makes essential use of distributivity and changes of variables, both for the first identity and the last.

Example 1.14. The set $\{0\}$, with $1 = 0$, and addition and multiplication given by $0 + 0 = 0 \times 0 = 0$ is a ring. We call it the **trivial ring**[1].

We call a ring in which $1 \neq 0$ a **non-trivial** ring.

Proposition 1.15. *Suppose that R is not a non-trivial ring. Then R is the trivial ring.*

Proof. Since R is not non-trivial, $1 = 0$ and hence for $x \in R$ we have $x = 1x = 0x = 0$ since zero annihilates. It follows that $R = \{0\}$ and there is only one function into a set of size one, so the addition and multiplication are uniquely determined and must be that of the trivial ring. □

Remark 1.16. In view of this proposition a ring is non-trivial if and only if it is not the trivial ring.

Example 1.17. The set of integers, $\mathbb{Z}$, is a ring with its usual addition and multiplication, zero and multiplicative identity

Remark 1.18. We write $\mathbb{N}^*$ for the positive integers, that is $\{1, 2, 3, \ldots\}$, and $\mathbb{N}_0$ for the non-negative integers, that is $\{0, 1, 2, \ldots\}$.

[1] Some authors (*e.g.* [Lam07]) use the term **zero ring**.

Ring homomorphisms

A map $\phi : R \to S$ between two rings is called **additive** if it is a homomorphism of the additive groups, and **multiplicative** if $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in R$. We say that ϕ is a **ring homomorphism** if it is additive and multiplicative and has $\phi(1_R) = 1_S$.

Remark 1.19. ⚠ If R is a non-trivial ring (for example $\mathbb{Z}$) and $\phi : \{0\} \to R$ is defined by $\phi(0) = 0$, then ϕ is additive and multiplicative but $0 = 1$ in the domain so $\phi(1) = \phi(0) = 0 \neq 1$ since R is non-trivial. Thus ϕ is *not* a ring homomorphism. In particular the condition that $\phi(1) = 1$ may *not* be dropped in the definition of ring homomorphism.

Example 1.20. For R a ring the identity map $\iota_R : R \to R; x \mapsto x$ is a ring homomorphism.

Example 1.21. The **zero map** $z_R : R \to \{0\}; x \mapsto 0$ from a ring R to the trivial ring is a ring homomorphism.

Remark 1.22. Since a ring homomorphism $\phi : R \to S$ is a homomorphism between the additive groups, $\phi(0_R) = 0_S$ and $\phi(-x) = -\phi(x)$ for all $x \in R$.

Lemma 1.23. *Suppose that $\phi : R \to S$ and $\psi : S \to T$ are ring homomorphisms. Then $\psi \circ \phi$ is a ring homomorphism $R \to T$.*

Proof. This is immediate from the definition. □

Theorem 1.24. *Suppose that R is a ring. Then there is a unique ring homomorphism $\chi_R : \mathbb{Z} \to R$.*

Remark 1.25. We omit the proof though it is not difficult: the idea is to define χ_R recursively on $\mathbb{N}^*$, first by $\chi_R(n) := 1_R + \cdots + 1_R$, where the sum is n-fold, and then extend this to $\mathbb{Z}$ by putting $\chi_R(n - m) := \chi_R(n) - \chi_R(m)$ for $n, m \in \mathbb{N}^*$.

Remark 1.26. The **characteristic** of a ring R is the smallest $n \in \mathbb{N}^*$ such that $\chi_R(n) = 0$ if such an n exists, and otherwise the characteristic is 0.

Example 1.27. The integers have characteristic 0.

Isomorphisms

A **ring isomorphism** is a map $\phi : R \to S$ that is a bijective ring homomorphism. We write $R \cong S$ if there is a ring isomorphism $R \to S$.

Example 1.28. Any ring Z enjoying the conclusions of Theorem 1.24 is isomorphic to $\mathbb{Z}$: For such a Z there is a ring homomorphism $\psi : Z \to \mathbb{Z}$, and by Theorem 1.24 there is a ring homomorphism $\phi : \mathbb{Z} \to Z$. By Lemma 1.23, the maps $\psi \circ \phi : \mathbb{Z} \to \mathbb{Z}$ and $\phi \circ \psi : Z \to Z$ are ring homomorphisms. So are the identity maps $\iota_{\mathbb{Z}} : \mathbb{Z} \to \mathbb{Z}$ and $\iota_Z : Z \to Z$, so by the

uniqueness part of the conclusion of Theorem 1.24 it follows that $\psi \circ \phi = \iota_{\mathbb{Z}}$, and by the uniqueness part of the hypothesis on Z that $\phi \circ \psi = \iota_Z$. The former identity tells us that ψ is a surjection, and the latter that ψ is an injection. In other words the conclusions of Theorem 1.24 characterise the integers up to isomorphism.[2]

Lemma 1.29. *Suppose that $\phi : R \to S$ is a ring isomorphism. Then ϕ^{-1} is a ring homomorphism, and hence a ring isomorphism.*

Proof. First, $\phi(1) = 1$ and ϕ is a bijection so $\phi^{-1}(1) = 1$. Secondly, ϕ is a bijective group homomorphism between the additive groups of R and S, and so ϕ^{-1} is also a group homomorphism between the additive groups of R and S. Finally, if $x, y \in S$ then by surjectivity there are elements $u, v \in R$ such that $\phi(u) = x$ and $\phi(v) = y$, and $\phi^{-1}(xy) = \phi^{-1}(\phi(u)\phi(v)) = \phi^{-1}(\phi(uv)) = uv = \phi^{-1}(x)\phi^{-1}(y)$. The result is proved. □

Proposition 1.30. $\cong$ *is an equivalence relation.*

Proof. The identity map on a ring is an isomorphism so $\cong$ is reflexive. $\cong$ is symmetric in view of Lemma 1.29. Finally, $\cong$ is transitive since the composition of bijections is a bijection, and composition of ring homomorphisms is a ring homomorphism – this is Lemma 1.23. □

Subrings

A ring S is a **subring** of a ring R if the map $j : S \to R; s \mapsto s$ is a well-defined[3] ring homomorphism called the **inclusion homomorphism**; S is **proper** if $S \neq R$.

Remark 1.31. ⚠ $\{0\}$ is a subset of $\mathbb{Z}$, and the operations on $\mathbb{Z}$ restrict to operations on $\{0\}$ giving it the structure of a ring (in fact the trivial ring), however $\{0\}$ is *not* a subring of $\mathbb{Z}$ since the inclusion map is not a ring homomorphism as shown in Remark 1.19.

Proposition 1.32 (Subring test)**.** *Suppose that R is a ring and $S \subset R$ has $1 \in S$ and $x - y, xy \in S$ for all $x, y \in S$. Then the addition and multiplication on R restrict to well-defined operations on S giving it the structure of a subring of R.*

Proof. First S is non-empty and $x - y \in S$ whenever $x, y \in S$ so by the subgroup test addition on R restricts to a well-defined binary operation on S giving it the structure of a commutative group.

Since S is closed under multiplication (meaning $xy \in S$ whenever $x, y \in S$) it also restricts to a well-defined binary operation on S, and is *a fortiori* associative since multiplication is associative on R. By hypothesis $1 \in S$ and since this is an identity for R it is *a fortiori* an

[2] One might describe the integers as one ring (up to isomorphism) ruling (uniquely embedding in) all others. [Tol04, Book I, Chapter 2, p66]

[3] All this does is ensure that $S \subset R$.

identity for S. Finally, multiplication and addition so restricted are *a fortiori* distributive on S. We conclude that S with these restricted operations is a ring.

The map $S \to R; s \mapsto s$ is then well-defined since S is a subset of R, and a ring homomorphism when S is equipped with these restricted operations, which is to say S so equipped is a subring of R. □

Remark 1.33. Given a subset satisfying the hypotheses of the above lemma, we make the common abuse of calling it a subring on the understanding that we are referring to the restricted operations described by the lemma.

Example 1.34. The set $\mathbb{N}^*$ contains 1 and if $x, y \in \mathbb{N}^*$ then $x + y, xy \in \mathbb{N}^*$, but $\mathbb{N}^*$ is *not* a subring of $\mathbb{Z}$ because it does not contain 0. It follows that $x - y$ may not be replaced by $x + y$ in the hypotheses of the subring test.

Example 1.35. $\mathbb{Z}$ has no proper subrings: Given a subring R of $\mathbb{Z}$, by Theorem 1.24 there is a ring homomorphism $\mathbb{Z} \to R$; composition with the inclusion homomorphism $R \to \mathbb{Z}$ gives a ring homomorphism $\mathbb{Z} \to \mathbb{Z}$. By Theorem 1.24 this map must be the identity map, and so the inclusion homomorphism $R \to \mathbb{Z}$ must be surjective and $R = \mathbb{Z}$ as claimed.

Fields and vector space structures

Write[4] R^* for the set of non-zero elements of a ring R. We say that a commutative ring R is a **field** if multiplication on R restricts to a binary operation on R^* making it into a group with identity 1.

Remark 1.36. If a subring of a ring is also a field we call it a **subfield**; it is **proper** if it is a proper subring.

Example 1.37. The field of **rationals**, $\mathbb{Q}$, is a field with $\mathbb{Z}$ as a subring such that every rational can be written in the form z/w for some $z \in \mathbb{Z}$ and $w \in \mathbb{Z}^*$.

The construction of the rationals from the integers is part of a general construction of fields of fractions which we shall meet later in Theorem 2.16.

Example 1.38. The field of **reals**, $\mathbb{R}$, is a field containing $\mathbb{Q}$ as a subfield and with a strict total order $>$ that is compatible with addition in the sense that $z + x > z + y$ whenever $x > y$, and multiplication in the sense that $xy > 0$ whenever $x, y > 0$, and satisfying the completeness axiom, meaning that if $S \subset \mathbb{R}$ is a non-empty set that is bounded above then S has a supremum.

[4] ⚠ Some authors (*e.g.* [Lan02, p84] and [Lam07, xiv]) write R^* for the group of units of R which we shall define later.

It turns out that ring homomorphisms from fields induce vector space structures on their codomain:

Proposition 1.39. *Suppose that $\phi : \mathbb{F} \to R$ is a ring homomorphism from a field $\mathbb{F}$. Then the map $. : \mathbb{F} \times R \to R$ defined by $\lambda.v := \phi(\lambda)v$ gives R the structure of an $\mathbb{F}$-vector space so that right multiplication in R is $\mathbb{F}$-linear.*

Proof. First, $1_{\mathbb{F}}.v = \phi(1_{\mathbb{F}})v = 1_R v = v$ since 1_R is an identity. Since multiplication on R is associative and ϕ is multiplicative we have $\lambda.(\mu.v) = \phi(\lambda)(\phi(\mu)v) = (\phi(\lambda)\phi(\mu))v = \phi(\lambda\mu)v = (\lambda\mu).v$. Since ϕ is a homomorphism of the additive group of $\mathbb{F}$ and right multiplication by v is a group homomorphism of the additive group of R we have $(\lambda + \mu).v = \phi(\lambda + \mu)v = (\phi(\lambda) + \phi(\mu))v = \phi(\lambda)v + \phi(\mu)v = \lambda.v + \mu.v$. Since left multiplication by $\phi(\lambda)$ is a group homomorphism of the additive group of R we have $\lambda.(v+w) = \phi(\lambda)(v+w) = \phi(\lambda)v+\phi(\lambda)w = \lambda.v+\lambda.w$. Finally, $\lambda.(vw) = \phi(\lambda)(vw) = (\phi(\lambda)v)w = (\lambda.v)w$ since multiplication is associative so right multiplication is $\mathbb{F}$-linear and the result is proved. □

Remark 1.40. We call the vector space structure of this proposition the **$\mathbb{F}$-vector space structure induced by ϕ**.

The hypothesis that $\phi : \mathbb{F} \to R$ is a ring homomorphism from a field is more restrictive than one might at first suppose:

Proposition 1.41. *Suppose that $\phi : \mathbb{F} \to R$ is a ring homomorphism, $\mathbb{F}$ is a field and R is non-trivial. Then ϕ is injective.*

Proof. If $\phi(x) = \phi(y)$ and $x \neq y$ then $x-y \in \mathbb{F}^*$ and so there is u such that $(x-y)u = 1$ whence $0 = 0\phi(u) = (\phi(x) - \phi(y))\phi(u) = \phi((x-y)u) = \phi(1) = 1$, which contradicts the non-triviality of R. The result is proved. □

Example 1.42. $\mathbb{Q}$ is a subfield of $\mathbb{R}$, and the $\mathbb{Q}$-vector space structure induced by the inclusion homomorphism then makes $\mathbb{R}$ into a vector space such that multiplication in $\mathbb{R}$ is $\mathbb{Q}$-bilinear. (Bilinear rather than just linear in the first argument since multiplication in $\mathbb{R}$ is commutative.)

Example 1.43. The field of **complex numbers**, $\mathbb{C}$, is a field with $\mathbb{R}$ as a subfield, such that 1 and i are a basis for $\mathbb{C}$ in the $\mathbb{R}$-vector space structure induced the inclusion homomorphism $\mathbb{R} \to \mathbb{C}$, and where the multiplication on $\mathbb{C}$ is determined by $i^2 = -1$. In particular, for $z, w \in \mathbb{C}$ there are unique $x, y, u, v \in \mathbb{R}$ such that $z = x + iy$ and $w = u + iv$, and we have

$$z + w = (x + u) + i(y + v), \text{ and } zw = (xu - yv) + i(xv + yu).$$

Example 1.44. Complex conjugation is a ring isomorphism from $\mathbb{C}$ to itself: it is a ring homomorphism since $\overline{1} = 1$; $\overline{z + w} = \overline{z} + \overline{w}$; and $\overline{zw} = \overline{z}\,\overline{w}$; it is a bijection since $\overline{\overline{z}} = z$ which is to say complex conjugation is self-inverse.

Example 1.45. The ring $\mathbb{Z}[i] \coloneqq \{a+bi : a,b \in \mathbb{Z}\}$ is called the ring of **Gaussian integers**. It is a subring of $\mathbb{C}$ by the subring test since $1 \in \mathbb{Z}[i]$ and if $x,y \in \mathbb{Z}[i]$ then $x = a+bi$ and $y = c+di$ for $a,b,c,d \in \mathbb{Z}$, and $x-y = (a-c)+(b-d)i \in \mathbb{Z}[i]$, while $xy = (ac-bd)+(ad+bc)i \in \mathbb{Z}[i]$.

Prototypical rings

Groups of symmetries are the prototypes for abstract groups and rings have a similar prototype which we shall now describe.

Remark 1.46. For a commutative group M we denote the group operation by $+$, the additive inverse of x by $-x$, and the identity by 0. We use subscripts to distinguish operations on different groups, for example writing $+_M$ and 0_M for the operation and identity on the commutative group M. This notation agrees with the notation for the additive group of a ring.

Remark 1.47. For N a commutative group, the set of maps $X \to N$ is a commutative group under pointwise addition: $(f+g)(x) = f(x)+g(x)$ for all $x \in X$. The first $+$ here is the addition on the set of functions $X \to N$, and the second is the addition on N. The identity is the map $z : X \to N; x \mapsto 0_N$, and the additive inverse of $\phi : X \to N$ is the map $X \to N; x \mapsto -\phi(x)$.

Lemma 1.48. *Suppose that M and N are commutative groups. Then* $\mathrm{Hom}(M,N)$*, the set of group homomorphisms $M \to N$, is itself a commutative group under pointwise addition.*

Suppose P is a further commutative group and $\phi \in \mathrm{Hom}(M,N)$ and $\psi \in \mathrm{Hom}(N,P)$, then $\psi \circ \phi \in \mathrm{Hom}(M,P)$ where $\circ$ denotes composition of functions; if $\pi \in \mathrm{Hom}(M,N)$ then $\psi \circ (\phi +_N \pi) = (\psi \circ \phi) +_P (\psi \circ \pi)$; and if $\pi \in \mathrm{Hom}(N,P)$ then $(\psi +_P \pi) \circ \phi = (\psi \circ \phi) +_P (\pi \circ \phi)$.

Proof. Suppose that $\phi, \psi \in \mathrm{Hom}(M,N)$. Then for all $x,y \in M$ we have

$$\begin{aligned}
(\phi +_N \psi)(x +_M y) &= \phi(x +_M y) +_N \psi(x +_M y) \\
&= (\phi(x) +_N \phi(y)) +_N (\psi(x) +_N \psi(y)) && \text{ϕ and ψ are group homomorphisms} \\
&= (\phi(x) +_N \psi(x)) +_N (\phi(y) +_N \psi(y)) && \text{associativity and commutativity of } +_N \\
&= (\phi +_N \psi)(x) +_N (\phi +_N \psi)(y). && \text{definition of pointwise addition}
\end{aligned}$$

It follows that $\phi +_N \psi \in \mathrm{Hom}(M,N)$. The map z is a homomorphism because $z(x)+z(y) = 0_N + 0_N = 0_N = z(x+y)$, and if $\phi \in \mathrm{Hom}(M,N)$ then the map $M \to N; x \mapsto -\phi(x)$ is a homomorphism because it is the composition of the homomorphism ϕ and negation which is a homomorphism on N since $+_N$ is commutative. By the subgroup test $\mathrm{Hom}(M,N)$ is a subgroup.

For the second part of the proposition recall that the composition of homomorphisms is a homomorphism which says exactly that if $\phi \in \mathrm{Hom}(M,N)$ and $\psi \in \mathrm{Hom}(N,P)$, then

$\psi \circ \phi \in \mathrm{Hom}(M,P)$. Now, if $\phi, \pi \in \mathrm{Hom}(M,N)$ and $\psi \in \mathrm{Hom}(N,P)$, then

$$\psi \circ (\phi +_N \pi)(x) = \psi(\phi(x) +_N \pi(x)) = \psi(\phi(x)) +_P \psi(\pi(x)) = ((\psi \circ \phi) +_P (\psi \circ \pi))(x)$$

by definition and the fact that ψ is a homomorphism, and we have that $\psi \circ (\phi +_N \pi) = (\psi \circ \phi) +_P (\psi \circ \pi)$ as claimed. On the other hand, if $\phi \in \mathrm{Hom}(M,N)$ and $\psi, \pi \in \mathrm{Hom}(N,P)$, then

$$(\psi +_P \pi) \circ \phi(x) = \psi(\phi(x)) +_P \pi(\phi(x)) = ((\psi \circ \phi) +_P (\pi \circ \phi))(x)$$

by definition. The result is proved. □

Remark 1.49. To show that $\mathrm{Hom}(M,N)$ is a subgroup it is essential that N be a *commutative* group, not just any group.

Remark 1.50. For the identity $\psi \circ (\phi +_N \pi) = (\psi \circ \phi) +_P (\psi \circ \pi)$ we used the homomorphism property of ψ, while the identity $(\psi +_P \pi) \circ \phi = (\psi \circ \phi) +_P (\pi \circ \phi)$ followed simply from the definition; *c.f.* Exercise I.2.

Theorem 1.51. *Suppose that M is a commutative group. Then the set $\mathrm{Hom}(M,M)$ equipped with pointwise addition as its addition and functional composition as its multiplication is a ring whose multiplicative identity is the map $M \to M; x \mapsto x$.*

Proof. By the first part of Lemma 1.48 $\mathrm{Hom}(M,M)$ is a commutative group under this addition, and by the second part the proposed multiplication distributes over this addition. It remains to recall that composition of functions is associative so the proposed multiplication is associative, and the map $M \to M; x \mapsto x$ is certainly a homomorphism and an identity for composition. □

Theorem 1.52. *Suppose that R is a ring. The map $\Psi : R \to \mathrm{Hom}(R,R); r \mapsto (R \to R; x \mapsto rx)$ is an injective ring homomorphism.*

Proof. The R in $\mathrm{Hom}(R,R)$ is the additive group of R. Ψ is multiplicative since $\Psi(rs)(x) = (rs)(x) = r(s(x)) = \Psi(r) \circ \Psi(s)(x)$; Ψ is additive since $\Psi(r+s) = (r+s)x = rx + sx = (\Psi(r) + \Psi(s))(x)$; and finally $\Psi(1_R)(x) = 1Rx = x = 1_{\mathrm{Hom}(R,R)}(x)$, so Ψ is a ring homomorphism. Ψ is injective because if $\Psi(r) = \Psi(s)$ then $s = s1_R = \Psi(s)(1_R) = \Psi(r)(1_R) = r1_R = r$. □

Remark 1.53. This can be thought of as 'Cayley's Theorem for rings'.

Remark 1.54. $\chi_{\mathrm{Hom}(R,R)} = \Psi \circ \chi_R$ by Theorem 1.24, and since Ψ is injective, the characteristic of R is the same as that of $\mathrm{Hom}(R,R)$.

Products of rings

Proposition 1.55. *Suppose that $R_1, \dots, R_n$ are rings. Then the product group $R_1 \times \cdots \times R_n$ of the additive groups of the rings R_i may be equipped with the structure of a ring with multiplication defined by $(xy)_i := x_i y_i$ for all $1 \leqslant i \leqslant n$ and, and identity $1 = (1_{R_1}, \dots, 1_{R_n})$. Moreover, $R_1 \times \cdots \times R_n$ is commutative iff R_i is commutative for all $1 \leqslant i \leqslant n$.*

Proof. The product group $R_1 \times \cdots \times R_n$ is a commutative group since the R_is are commutative groups. Associativity of multiplication follows coordinate-wise from associativity of multiplication in each of the rings R_i: for x, y, z we have $(x(yz))_i = x_i(yz)_i = x_i(y_i z_i) = (x_i y_i)z_i = (xy)_i z_i = ((xy)z)_i$. Similarly, the fact that left multiplication is a homomorphism follows since $(x(y+z))_i = x_i(y+z)_i = x_i(y_i + z_i) = (x_i y_i) + (x_i z_i) = (xy)_i + (xz)_i = (xy + xz)_i$, this time using the fact that addition in this group is coordinate-wise. The fact that right multiplication is a homomorphism follows similarly, and hence multiplication is distributive. Finally, $(1x)_i = 1_{R_i} x_i = x_i = x$ and similarly $x1 = x$.

Multiplication on R_i is commutative if and only if $(xy)_i = x_i y_i = y_i x_i = (yx)_i$ for all x, y, and so multiplication on $R_1 \times \cdots \times R_n$ is commutative if and only if it is commutative in R_i for all i. □

Remark 1.56. By the ring $R_1 \times \cdots \times R_n$ we shall mean the ring of Proposition 1.55 - it is called the **direct product** of the R_is.

Remark 1.57. By the ring R^n we mean the ring $R \times \cdots \times R$ with R occurring n times.

Example 1.58. The ring $\mathbb{R}^2$ is *not* ring isomorphic to the the ring $\mathbb{C}$. In particular $(0,1), (1,0) \in (\mathbb{R}^2)^*$ have $(0,1) \times (0,1) = (0,0)$, but if $x, y \in \mathbb{C}^*$ then $xy \neq 0$ and so there can be no ring isomorphism between $\mathbb{R}^2$ and $\mathbb{C}$.

This is despite the fact that the additive groups are isomorphic as groups, and even more the $\mathbb{R}$-vector space structures on $\mathbb{R}^2$ induced by the map $\mathbb{R} \to \mathbb{R}^2; \lambda \mapsto (\lambda, \lambda)$ and on $\mathbb{C}$ induced by the inclusion homomorphism $\mathbb{R} \to \mathbb{C}$, are isomorphic.

Matrix rings

Given a ring R we write $M_{n,m}(R)$ for the set of $n \times m$ matrices with values in R.

Remark 1.59. $M_{n,m}(R)$ is a commutative group with addition defined by

$$A + B := (A_{i,j} + B_{i,j})_{i=1,j=1}^{n,m} \text{ for all } A, B \in M_{n,m}(R).$$

The additive inverse of $A \in M_{n,m}(R)$ is $-A = (-A_{i,j})_{i=1,j=1}^{n,m}$, and the identity is $0_{n \times m}$, the matrix with 0_R in every entry.

For $A \in M_{n,m}(R)$ and $B \in M_{m,p}(R)$ then we define a matrix $AB \in M_{n,p}(R)$ by

$$(AB)_{i,k} \coloneqq \sum_{j=1}^{m} A_{i,j}B_{j,k}. \tag{1.1}$$

We write $M_n(R) \coloneqq M_{n,n}(R)$ and the matrix I denotes the matrix with 1_Rs on the diagonal and 0_Rs elsewhere *i.e.* $I_{i,i} = 1_R$ for all $1 \leqslant i \leqslant n$ and $I_{i,j} = 0_R$ for all $i \neq j$. If we need to make it clear that I is $n \times n$ we write I_n.

Proposition 1.60 (Algebra of matrix multiplication)**.** *Suppose that R is a ring, $A \in M_{n,m}(R)$, $B, B' \in M_{m,l}(R)$, and $C, C' \in M_{p,n}(R)$. Then $C(AB) = (CA)B$, $A(B + B') = (AB) + (AB')$, $(C + C')A = (CA) + (C'A)$, $AI_m = A$ and $I_nA = A$.*

Proof. First, for $1 \leqslant i \leqslant p$ and $1 \leqslant j \leqslant l$ we have

$$\begin{aligned}
(C(AB))_{i,j} &= \sum_{k=1}^{n} C_{i,k}\left(\sum_{l=1}^{m} A_{k,l}B_{l,j}\right) \\
&= \sum_{k=1}^{n}\left(\sum_{l=1}^{m} C_{i,k}(A_{k,l}B_{l,j})\right) \\
&= \sum_{k=1}^{n}\left(\sum_{l=1}^{m} (C_{i,k}A_{k,l})B_{l,j}\right) \\
&= \sum_{l=1}^{m}\left(\sum_{k=1}^{n} (C_{i,k}A_{k,l})B_{l,j}\right) \\
&= \sum_{l=1}^{m}\left(\sum_{k=1}^{n} C_{i,k}A_{k,l}\right)B_{l,j} = ((CA)B)_{i,j}.
\end{aligned}$$

left multiplication by $C_{i,k}$ is a homomorphism of the additive group

associativity of multiplication

change of variables (Remark 1.7)

right multiplication by $B_{l,j}$ is a homomorphism of the additive group

This gives the first identity. For $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant l$ we have

$$\begin{aligned}
(A(B + B'))_{i,j} &= \sum_{k=1}^{n} A_{i,k}(B_{k,j} + B'_{k,j}) \\
&= \sum_{k=1}^{n} (A_{i,k}B_{k,j} + A_{i,k}B'_{k,j}) \\
&= \left(\sum_{k=1}^{n} A_{i,k}B_{k,j}\right) + \left(\sum_{k=1}^{n} A_{i,k}B'_{k,j}\right) = (AB)_{i,j} + (AB')_{i,j}.
\end{aligned}$$

left multiplication by $A_{i,k}$ is a homomorphism of the additive group

change of variables (Remark 1.7)

This gives the second identity, and the third follows in a similar way. Finally, $(AI)_{i,j} = \sum_{k=1}^{n} A_{i,k}I_{k,j} = 0_R + \cdots + 0_R + A_{i,j}1_R + 0_R + \cdots + 0_R = A_{i,j}$ since zero annihilates, 1_R is an identity for R, and sums of zeros are zero. Hence $AI_m = A$, and similarly $I_nA = A$. □

Remark 1.61. We identify R with $M_{1,1}(R)$, often writing λ in place of (λ), and R^n with $M_{1,n}(R)$, often writing $(a_1, \ldots, a_n)$ in place of $\begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix}$ ⚠ I have a natural inclination when writing matrices to take R^n to mean $M_{n,1}(R)$. I have tried to eliminate this in these notes.

Corollary 1.62. *Suppose that R is a ring. Then the commutative group $M_n(R)$ is a ring when endowed with the multiplication $(A, B) \mapsto AB$ with multiplicative identity I.*

Proof. By design the proposed multiplication is a binary operation on the commutative group $M_n(R)$. By Proposition 1.60 it is associative, distributes over the addition, and has an identity matrix I. The result is proved. □

Example 1.63. For R non-trivial the ring $M_2(R)$ is not commutative:

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Example 1.64. The map $R \to M_n(R); \lambda \mapsto \lambda I$ is a ring homomorphism. If R is commutative then everything in the image of R commutes with the image of this homomorphism, meaning $(\lambda I)A = A(\lambda I)$ for all $A \in M_n(R)$ and $\lambda \in R$ - this will be useful for applying polynomials to matrices in the forthcoming Proposition 1.73. If $R = \mathbb{F}$ is a field this induces the usual $\mathbb{F}$-vector space structure on $M_n(\mathbb{F})$.

Polynomial rings

Theorem 1.65. *Suppose that R is a non-trivial commutative ring. Then there is a ring $R[X]$ is a non-trivial commutative ring with R as a subring, and a distinguished element $X \in R[X]$ such that*

$$R[X] = \{a_0 + a_1X + \cdots + a_nX^n : a_0, \ldots, a_n \in R\}, \tag{1.2}$$

and

$$a_0 + a_1X + \cdots + a_nX^n = 0_R \Rightarrow a_0, \ldots, a_n = 0_R. \tag{1.3}$$

Remark 1.66. We shall not show that a ring with these properties exists though it is not difficult to do so.

Remark 1.67. The ring $R[X]$ from this theorem is called the **polynomial ring over** R **with indeterminate** X.

Remark 1.68. Since R is a subring of $R[X]$, $R[X]$ has zero and multiplicative identity of 0_R and 1_R respectively.

Remark 1.69. Given a field $\mathbb{F}$ there is an $\mathbb{F}$-vector space structure induced on $\mathbb{F}[X]$ by the inclusion homomorphism $\mathbb{F} \to \mathbb{F}[X]$ (Proposition 1.39). In this structure (1.2) says exactly that $\{1, X, X^2, \ldots\}$ is a spanning set, while (1.3) tells us it is linearly independent; together this means it forms a basis.

Remark 1.70. If $a_0 + a_1X + \cdots + a_nX^n = b_0 + b_1X + \cdots + b_mX^m$ for $a_0, a_1, \ldots, b_0, b_1, \cdots \in R$ with $a_i = 0$ for $i > n$ and $b_j = 0$ for $j > m$, then $(a_0 - b_0) + (a_1 - b_1)X + \cdots + (a_k - b_k)X^k = 0$ for $k = \max\{n, m\}$ by Proposition 1.12. From (1.3) we conclude that $a_i = b_i$ for all i - this inference justifies the process of 'equating coefficients' between polynomials.

Remark 1.71. For $p \in R[X]^*$ there is a minimal $d \in \mathbb{N}_0$ such that there are $a_0, a_1, \ldots, a_d \in R$ with $p(X) = a_0 + a_1X + \cdots + a_dX^d$. Since zero annihilates and sums of zeros are zero, the reverse implication in (1.3) holds and so $a_d \neq 0_R$.

We call this minimal d the **degree** of p and denote it $\deg p$; we call a_i the **coefficient** of X^i; a_d the **lead coefficient** and a_0 the **constant coefficient**. The use of 'the' here is justified by equating coefficients as discussed in Remark 1.70.

The **constant polynomials** are those for which the constant coefficient is the only coefficient that may be non-zero. The **monic** polynomials are those where the lead coefficient is 1.

Remark 1.72. We define $R[X_1, \ldots, X_n] \coloneqq R[X_1, \ldots, X_{n-1}][X_n]$ and call $R[X_1, \ldots, X_n]$ the **polynomial ring in the indeterminates** $X_1, \ldots, X_n$.

Proposition 1.73. *Suppose that $\phi : R \to S$ is a ring homomorphism from a commutative ring R, and $\lambda \in S$ commutes with all elements in the image of ϕ, meaning $\phi(r)\lambda = \lambda\phi(r)$ for all $r \in R$. Then there is a unique homomorphism $\widetilde{\phi} : R[X] \to S$ such that $\widetilde{\phi}(r) = \phi(r)$ for all $r \in R$ and $\widetilde{\phi}(X) = \lambda$.*

Proof. We begin with existence: Define $\widetilde{\phi}$ by

$$\widetilde{\phi}(a_0 + a_1X + \cdots + a_nX^n) \coloneqq \phi(a_0) + \phi(a_1)\lambda + \cdots + \phi(x_n)\lambda^n. \tag{1.4}$$

This is well-defined by (1.2) and (1.3). $\widetilde{\phi}(1_{R[X]}) = 1_S$ since $\phi(1_R) = 1_S$. $\widetilde{\phi}$ is additive in view of Proposition 1.12 and the fact that ϕ is additive. For $b_0, \ldots, b_m, a_0, \ldots, a_n \in R$, b_i commutes with X for all i since $R[X]$ is commutative so by Proposition 1.12 we have

$$\widetilde{\phi}\left(\left(\sum_{i=0}^{n} a_iX^i\right)\left(\sum_{j=0}^{m} b_jX^j\right)\right) = \sum_{k=0}^{n+m} \phi\left(\sum_{j=0}^{k} a_{k-j}b_j\right)\lambda^k, \tag{1.5}$$

and $\phi(b_i)$ commutes with λ for all i by hypothesis, so by Proposition 1.12 we have

$$\widetilde{\phi}\left(\sum_{i=0}^{n} a_iX^i\right)\widetilde{\phi}\left(\sum_{j=0}^{m} b_jX^j\right) = \left(\sum_{i=0}^{n} \phi(a_i)\lambda^i\right)\left(\sum_{j=0}^{m} \phi(b_j)\lambda^j\right) = \sum_{k=0}^{n+m}\left(\sum_{j=0}^{k} \phi(a_{k-j})\phi(b_j)\right)\lambda^k. \tag{1.6}$$

Since ϕ is additive and multiplicative we have

$$\phi\left(\sum_{j=0}^{k} a_{k-j}b_j\right) = \left(\sum_{j=0}^{k} \phi(a_{k-j})\phi(b_j)\right),$$

and so the right hand sides of (1.5) and (1.6) are equal, and hence $\widetilde{\phi}$ is multiplicative. Finally from (1.4), $\widetilde{\phi}(r) = \phi(r)$ for all $r \in R$ and $\widetilde{\phi}(X) = \widetilde{\phi}(1X) = \phi(1)\lambda = 1\lambda = \lambda$ as required.

Uniqueness follows because if $\widetilde{\phi} : R[X] \to S$ is a ring homomorphism with $\widetilde{\phi}(X) = \lambda$ and $\widetilde{\phi}(r) = \phi(r)$ for all $r \in R$ then (1.4) must hold. □

Remark 1.74. The homomorphism of this proposition is called the **evaluation homomorphism**. Sometimes the homomorphism $\phi : R \to S$ is implicit (for example it will often be inclusion) and in this case we write $p(\lambda)$ in place of $\widetilde{\phi}(p)$, and $R[\lambda]$ for the image of $\widetilde{\phi}$.

For example, if $A \in M_n(\mathbb{C})$ then $\mathbb{R}[A]$ denotes the set of matrices often written in the form $a_0 I + a_1 A + \cdots + a_n A^n$ for $a_0, \ldots, a_n \in \mathbb{R}$, and the implicit ϕ is the map $\mathbb{R} \to M_n(\mathbb{C}); \lambda \mapsto \lambda I$ which is the composition of the inclusion homomorphism $\mathbb{R} \to \mathbb{C}$ and the homomorphism $\mathbb{C} \to M_n(\mathbb{C})$ in Example 1.64.

The notation $R[X]$ (from Theorem 1.65) and $\mathbb{Z}[i]$ (from Example 1.45) can be consistently interpreted as arising through such homomorphisms: for the first, it is the inclusion homomorphism $R \to R[X]$, and for the second the inclusion homomorphism $\mathbb{Z} \to \mathbb{C}$.

2 Units and integral domains

An element $x \in R$ is a **unit** if there is some $y \in R$ such that $xy = yx = 1$; we write[4] $U(R)$ for the set of units of R.

Proposition 2.1. *Suppose that R is a ring. Then multiplication on R restricts to a well-defined binary operation on $U(R)$ giving it the structure of a group with identity 1. Furthermore, if R is commutative then so is this group, and if $\phi : R \to S$ is a ring homomorphism then $U(R) \to U(S); x \mapsto \phi(x)$ is a well-defined group homomorphism.*

Proof. First, suppose that $x, y \in U(R)$. Then there are elements $z, w \in R$ such that $xz = zx = 1$ and $yw = wy = 1$, so $(xy)(wz) = x((yw)z) = xz = 1$ and similarly $(wz)(xy) = 1$ so $xy \in U(R)$. Hence multiplication on R restricts to a well-defined binary operation on $U(R)$.

Since multiplication is associative on R, it is *a fortiori* associative when restricted to $U(R)$. Since 1 is an identity for $\times$ on R we have $1 \times 1 = 1$ and so $1 \in U(R)$, and it is *a fortiori* an identity for multiplication restricted to $U(R)$. Finally, if $x \in U(R)$ then there is $z \in R$ such that $xz = zx = 1$, but then $z \in U(R)$ and so every $x \in U(R)$ has an inverse w.r.t. multiplication restricted to $U(R)$.

For the last part, if R is commutative then multiplication is commutative on R and *a fortiori* it is commutative when restricted to $U(R)$; and if $x \in U(R)$ then there is $y \in R$ such that $xy = yx = 1$ so $\phi(x)\phi(y) = \phi(y)\phi(x) = \phi(1) = 1$ so $\phi(r) \in U(S)$ as required. The result is proved. □

Remark 2.2. We call $U(R)$ equipped with multiplication as above the **group of units** of the ring R, and if $x \in U(R)$ we write x^{-1} for the inverse of x in the group in the proposition.

Remark 2.3. $1^{-1} = 1$ since identities are self-inverse; and $(x^{-1})^{-1} = x$ for all $x \in U(R)$. In general inversion is not a homomorphism because multiplication is not commutative, but we do have $(xy)^{-1} = y^{-1}x^{-1}$.

Remark 2.4. If R is a finite commutative ring then $U(R)$ is a finite commutative group, but exactly which finite commutative groups occur as the group of units of a ring is an open problem called Fuchs' problem [Fuc58, Problem 72, p299].

Example 2.5. $U(\mathbb{F}) = \mathbb{F}^*$ for $\mathbb{F}$ a field. On the one hand, if $x \in \mathbb{F}^*$ then since multiplication restricted to $\mathbb{F}^*$ gives it the structure of a group with identity 1, there is $y \in \mathbb{F}^*$ such that $xy = yx = 1$ and hence $x \in U(\mathbb{F})$. On the other hand $0x = 0 \neq 1$ for all x so $0 \notin U(\mathbb{F})$, and hence if $x \in U(\mathbb{F})$ then $x \in \mathbb{F}^*$.

Example 2.6. $U(\mathbb{Z}) = \{-1, 1\}$. Indeed, suppose that $x, y \in \mathbb{Z}$ have $xy = 1$. If $x > 0$ then $y > 0$ and so $x, y \geqslant 1$. But then if $x > 1$ we have $1 = xy > y \geqslant 1$ a contradiction, so $x = 1$. Similarly if $x < 0$ then $x = -1$. Thus $U(\mathbb{Z}) \subset \{1, -1\}$. Moreover, both these are units.

Remark 2.7. In particular, $\mathbb{Z}$ is not a field since $U(\mathbb{Z}) = \{-1, 1\} \neq \mathbb{Z}^*$.

Example 2.8. $U(\mathbb{Z}[i]) = \{1, -1, i, -i\}$. Indeed, suppose that $x, y \in \mathbb{Z}[i]$ have $xy = 1$. Then there are integers a, b, c, d such that $x = a + bi$ and $y = c + di$. Taking absolute values of $(a + bi)(c + di) = 1$ we have $(a^2 + b^2)(c^2 + d^2) = 1$. We conclude $a^2 + b^2 = 1$, and hence $(a, b) \in \{(1, 0), (-1, 0), (0, 1), (0, -1)\}$ as claimed. Conversely, all the elements of $\{1, -1, i, -i\}$ are units.

Remark 2.9. We call $y \in R$ a **left** (resp. **right**) **zero-divisor** if the left (resp. right) multiplication-by-y map has a non-trivial kernel *i.e.* if there is some $x \neq 0$ such that $yx = 0$ (resp. $xy = 0$).

Example 2.10. 0_R is a zero-divisor if and only if R is non-trivial.

Example 2.11. The elements $(0, 1)$ and $(1, 0)$ in the product ring $\mathbb{R}^2$ have $(0, 1)(1, 0) = (0, 0)$, and so are both non-zero zero-divisors.

Remark 2.12. If x is a unit then left (resp. right) multiplication by x has left (resp. right) multiplication by x^{-1} as an inverse map. But left (resp. right) multiplication by x is a group homomorphism and so it is a group isomorphism. In particular, no unit is a zero-divisor.

Remark 2.13. A commutative ring R is an **integral domain** if it is non-trivial and R^* is closed under multiplication, meaning $xy \in R^*$ whenever $x, y \in R^*$.

Example 2.14. $\mathbb{Z}$ is an integral domain: First it is a non-trivial commutative ring. Secondly, if $x, y \neq 0$ then either $x, y > 0$ or $x, -y > 0$, or $-x, y > 0$, or $-x, -y > 0$. Since negation distributes, in the first and last case $xy > 0$ and so $xy \neq 0$; in the second and third case $-(xy) > 0$ and so $0 > xy$ and $xy \neq 0$. It follows that $\mathbb{Z}^*$ is closed under multiplication.

Remark 2.15. Suppose that R is a commutative ring. R is non-trivial if $1 \in R^*$ (since this exactly means $1 \neq 0$); R is an integral domain if $1 \in R^*$ and R^* is closed under multiplication; and R is a field if $1 \in R^*$, R^* is closed under multiplication, and has multiplicative inverses.

Theorem 2.16 (Field of fractions and its characterisation)**.** *Suppose that R is an integral domain. Then there is a field $\mathbb{F}$ with R as a subring, and no proper subfield of $\mathbb{F}$ contains R. Moreover, if $\mathbb{K}$ is a field with R as a subring, and no proper subfield of $\mathbb{K}$ contains R then there is a ring isomorphism $\phi : \mathbb{K} \to \mathbb{F}$ which is the identity on R.*

Remark 2.17. The proof is not hard and can be found in many places *e.g.* [Hun80, Theorem 4.3] and [Lan02, Chapter II, §4], but it is omitted from the syllabus. It is not dissimilar to the construction of the integers from the naturals by 'adding in' the negative numbers.

Remark 2.18. Suppose that $\mathbb{F}$ is a field of fractions for R, and consider the set $F(R) := \{ab^{-1} : a \in R, b \in R^*\}$ as a subset of $\mathbb{F}$. This contains $1 = 1.1^{-1}$ and is closed under subtraction and multiplication since

$$ac^{-1} - bd^{-1} = (ad - bc)(cd)^{-1} \text{ and } (ac^{-1})(bd^{-1}) = (ab)(cd)^{-1}.$$

It follows from the subring test that $F(R)$ is a subring and it contains R. Now, if $ab^{-1} \neq 0$ then $a \in R^*$ so $ba^{-1} \in F(R)$, and hence $F(R)$ is closed under multiplicative inverses and so a field, whence $F(R) = \mathbb{F}$. This motivates the name field of fractions: all the elements of $\mathbb{F}$ can be written as a 'fraction' ab^{-1}.

Example 2.19. The rationals $\mathbb{Q}$ are the field of fractions of the ring of integers $\mathbb{Z}$.

Proposition 2.20. *Suppose that R is a finite integral domain. Then R is a field.*

Proof. For $x \in R^*$ the map that is left multiplication by x is an injective homomorphism from the additive group to itself. Since R is finite, the additive group is finite and so every injection is a surjection. In particular there is $y \in R$ such that $xy = 1$, but R is commutative and hence $x \in U(R) = R^*$ as required. □

Remark 2.21. A finite ring R with no non-zero zero divisors is commutative, and hence a field. This is a result called Wedderburn's Little Theorem, a proof of which may be found in [Wit31].

Integral domains produce polynomial rings where the degree function behaves nicely:

Proposition 2.22. *Suppose that R is a non-trivial commutative ring. Then TFAE:*

(i) R is an integral domain;

(ii) $R[X]$ is an integral domain;

(iii) for every $p, q \in R[X]^*$ we have $pq \in R[X]^*$ and $\deg pq = \deg p + \deg q$.

Proof. Certainly (ii) implies (i) since R is a subring of $R[X]$, and (iii) implies (ii) since $R[X]$ is a non-trivial commutative ring, and so the fact it is an integral domain follows by forgetting the degree equation in (iii).

To see (i) implies (iii) suppose that $p, q \in R[X]^*$ have degree n and m, and lead coefficients a_n and b_m respectively. Then from Proposition 1.12 we see that $\deg pq \leqslant n + m$ and the coefficient of X^{n+m} is $a_n b_m$. The coefficient of X^{n+m} is non-zero since R is an integral domain and $a_n, b_m \in R^*$. We conclude that $pq \in R[X]^*$ and $\deg pq = n + m = \deg p + \deg q$ as required. □

Example 2.23. By Proposition 2.22 $\mathbb{Z}[X]$ is an integral domain since $\mathbb{Z}$ is an integral domain (Example 2.14)

Example 2.24. For $\mathbb{F}$ a field, $\mathbb{F}[X_1, \ldots, X_n]$ is an integral domain by induction on n: For $n = 0$, every field is an integral domain and we are done; then by definition (Remark 1.72) of $\mathbb{F}[X_1, \ldots, X_{n+1}] = \mathbb{F}[X_1, \ldots, X_n][X_{n+1}]$, so the left hand side is an integral domain by the inductive hypothesis and Proposition 2.22.

Proposition 2.25. *Suppose R is an integral domain. Then $U(R[X]) = U(R)$.*

Proof. Suppose that $p \in U(R[X])$. Then there is some $q \in U(R[X])$ such that $pq = 1$, and so by Proposition 2.22 (iii) we have $0 = \deg p + \deg q$ and so $\deg p = 0$ and $\deg q = 0$ and hence $p(X) = a_0$ and $q(X) = b_0$ for some $a_0, b_0 \in R^*$. Since $a_0 b_0 = 1$ and R is commutative we conclude that $a_0 \in U(R)$ as required. Conversely, if $p \in U(R)$ then $p \in U(R[X])$ and we are done. □

3 Ideals and quotient rings

We begin this section with an important and familiar example.

Example 3.1. For $N \in \mathbb{N}^*$, $N\mathbb{Z}$ – the set of integer multiples of N – is a normal subgroup of the additive group of the ring $\mathbb{Z}$. The quotient group $\mathbb{Z}/N\mathbb{Z}$ has its addition determined by the fact the quotient map $q : \mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}; x \mapsto x + N\mathbb{Z}$ is a homomorphism of groups, so $(x + N\mathbb{Z}) + (y + N\mathbb{Z}) = (x + y) + N\mathbb{Z}$.

It happens that the multiplication on $\mathbb{Z}$ also gives rise to a well-defined multiplication on $\mathbb{Z}/N\mathbb{Z}$ by $(x + N\mathbb{Z})(y + N\mathbb{Z}) = xy + N\mathbb{Z}$, which gives $\mathbb{Z}/N\mathbb{Z}$ the structure of a ring with multiplicative identity $1 + N\mathbb{Z}$. We denote this ring $\mathbb{Z}_N$, and write $x \equiv y \pmod N$ to mean $x + N\mathbb{Z} = y + N\mathbb{Z}$.

Not every subgroup of the additive group of a ring gives rise to a quotient group supporting a ring multiplication:

Example 3.2. The additive group of $\mathbb{Z}$ is a subgroup of the additive group of $\mathbb{Q}$, but the quotient group $\mathbb{Q}/\mathbb{Z}$ is *not* the additive group of any ring. To see this, write $S := \mathrm{Hom}(\mathbb{Q}/\mathbb{Z}, \mathbb{Q}/\mathbb{Z})$ and note that for $n \in \mathbb{N}^*$ we have $\chi_S(n)(1/2n + \mathbb{Z}) = 1/2 + \mathbb{Z} \neq \mathbb{Z} = 0_{\mathbb{Q}/\mathbb{Z}}$. Hence $\chi_S(n) \neq 0_S$ and the characteristic of S is not n. We conclude S has characteristic 0. If $\mathbb{Q}/\mathbb{Z}$ is the additive group of a ring R then $\mathrm{Hom}(R, R) = S$ and by Remark 1.54 it would have characteristic 0, but the multiplicative identity in this ring must be some element $a/b + \mathbb{Z}$ for $a \in \mathbb{Q}$ and $b \in \mathbb{N}^*$, and then $\chi_R(b) = (a/b + \mathbb{Z}) + \cdots + (a/b + \mathbb{Z}) = a + \mathbb{Z} = 0_R$ meaning R has non-zero characteristic; a contradiction.

Remark 3.3. To generalise Example 3.1 we need a suitable generalisation of 'multiples of N'. An **ideal**[5] I in a ring R is a subgroup of the additive group of R closed under left and right multiplication by elements of R, meaning $rx, xr \in I$ for all $x \in I$ and $r \in R$. The notation $I \lhd R$ is used in places (*e.g.* [Coh00, p12]) to mean I is an ideal of R.

Example 3.4. In any ring R the sets $\{0\}$ and R are ideals; the set $\{0\}$ is sometimes called the **zero ideal**.

Remark 3.5. ⚠ Note the difference between ideals and subrings: an ideal is closed under multiplication by any element of the containing ring, while a subring is only closed under multiplication by elements of itself. On the other hand a subring contains 1, while an ideal does not in general contain 1. The ring R itself is the only subset that is both an ideal and a subring.

Example 3.6 (Ideals in $\mathbb{Z}$)**.** For each $N \in \mathbb{N}_0$, $N\mathbb{Z}$ is an ideal in $\mathbb{Z}$. In fact all ideals in $\mathbb{Z}$ have this form: Suppose that I is a non-zero ideal in $\mathbb{Z}$ then I contains a positive element (since ideals are closed under multiplication by -1); let $N \in I$ be the minimal positive element of I. Of course $I \supset N\mathbb{Z}$; if $I \setminus N\mathbb{Z} \neq \varnothing$ then it contains a positive element and so a minimal positive element, say M. By minimality of N we have $M > N$ and of course $M - N \in I$. By minimality of M and positivity of $M - N$ we have $M - N \in N\mathbb{Z}$ whence $M \in N\mathbb{Z}$, a contradiction. It follows that $I = N\mathbb{Z}$ as claimed.

Example 3.7 (The fields $\mathbb{F}_p$)**.** If N is composite then $N = ab$ for $1 < a, b < N$ and so $ab \equiv 0 \pmod N$ while $a, b \not\equiv 0 \pmod N$. It follows that $\mathbb{Z}_N$ is *not* an integral domain.

On the other hand, if p is prime and $p \mid ab$ then $p \mid a$ or $p \mid b$, which in other language means if $ab \equiv 0 \pmod p$ then $a \equiv 0 \pmod p$ or $b \equiv 0 \pmod p$. It follows that $\mathbb{Z}_p$ is an integral domain and since it is finite it is a field by Proposition 2.20; we write $\mathbb{F}_p$ for $\mathbb{Z}_p$ to emphasise this property.[6]

[5] Also called a **two-sided ideal**.

[6] ⚠ $\mathbb{Z}_p$ is sometimes (*e.g.* [Lam07]) used to denote a different ring (which we shall not consider) called the **p-adic integers**.

Proposition 3.8. *Suppose that R is a ring and $I_1, \dots, I_n$ are ideals in R. Then $I_1 + \cdots + I_n$ is an ideal in R.*

Proof. Since the I_is are non-empty, the sum $I_1 + \cdots + I_n$ is non-empty. Suppose $x, y \in I_1 + \cdots + I_n$, then there are elements $x_i, y_i \in I_i$ such that $x = x_1 + \cdots + x_n$ and $y = y_1 + \cdots + y_n$. Hence $x - y = (x_1 - y_1) + \cdots + (x_n - y_n) \in I_1 + \cdots + I_n$. Moreover, if $r \in R$ then $rx = rx_1 + \cdots + rx_n \in I_1 + \cdots + I_n$ and $xr = x_1 r + \cdots + x_n r \in I_1 + \cdots + I_n$ and so $I_1 + \cdots + I_n$ is an ideal. □

Proposition 3.9. *Suppose that R is a ring and $(I_i)_{i \in X}$ is collection of ideals of R (with X non-empty). Then $\bigcap_{i \in X} I_i$ is an ideal.*

Proof. The requirement that X be non-empty ensures that the intersection is well-defined. Since I_i is an (additive) subgroup of R for all $i \in X$, we have $0_R \in I_i$ and hence $0_R \in \bigcap_{i \in X} I_i$. Now, suppose $x, y \in \bigcap_{i \in X} I_i$. Then $x, y \in I_i$ for all $i \in X$, and hence $x - y \in I_i$ for all $i \in X$, and $x - y \in \bigcap_{i \in X} I_i$; we conclude that $\bigcap_{i \in X} I_i$ is a subgroup by the subgroup test. Finally, if $x \in \bigcap_{i \in X} I_i$ and $r \in R$ then $x \in I_i$ for all $i \in X$, and hence $xr, rx \in I_i$ for all $i \in I$ so $xr, rx \in \bigcap_{i \in X} I_i$. The result is proved. □

Remark 3.10. Given a ring R and elements $x_1, \dots, x_n \in R$ we define

$$\langle x_1, \dots, x_n \rangle := \bigcap \{I : I \text{ is an ideal in } R \text{ and } x_1, \dots, x_n \in I\}, \tag{3.1}$$

which is an ideal by the preceding proposition (and Example 3.4 which ensures that R itself is an ideal so that some ideal contains $x_1, \dots, x_n$). We call $\langle x_1, \dots, x_n \rangle$ the **ideal generated by** $x_1, \dots, x_n$.

Remark 3.11. ⚠ The ideal generated by an element depends on the ambient ring: for example if $N \in \mathbb{N}^*$ then $\langle N \rangle = N\mathbb{Z}$ as an ideal in $\mathbb{Z}$, while $\langle N \rangle = \mathbb{Q}$ as an ideal in $\mathbb{Q}$.

Proposition 3.12. *Suppose that R is a commutative ring. Then $\langle x \rangle = Rx (= xR)$.*

Proof. $x \in \langle x \rangle$ and $\langle x \rangle$ is an ideal so $rx \in \langle x \rangle$ for all $r \in R$, whence Rx is contained in $\langle x \rangle$. On the other hand xR is the image of the additive group R under right multiplication by x. This map is a homomorphism and so this image – Rx – is a subgroup of the additive group of R. If $y \in Rx$ then there is $r \in R$ such that $y = rx$ and so if $s \in R$ then $sy = s(rx) = (sr)x \in Rx$, and since R is commutative $ys = sy \in Rx$. It follows that Rx is an ideal and it contains x so $\langle x \rangle \subset Rx$ and the result is proved. □

Example 3.13. ⚠ This is another place where we make use of the fact that our rings have multiplicative identities. The group $R = 2\mathbb{Z}$ with its usual addition and multiplication satisfies the axioms of a ring except that it does not have a multiplicative identity. The set $I := 6\mathbb{Z}$ is a subgroup of R which is closed under multiplication by elements of R (on the left and right), but there is no $x \in R$ such that $I = Rx$.

Example 3.14. Suppose that $\mathbb{F}$ is a field, $R = M_2(\mathbb{F})$ and

$$x := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \text{ and } p := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The set Rx is a set of matrices all of which have rank at most 1. By calculation I, the $2x2$ identity matrix, has $I = x + pxp$, and since $\langle x \rangle$ is an ideal it follows that $I \in \langle x \rangle$. But I has rank 2 and so is not in Rx, and we conclude that Rx is not an ideal. It follows that the commutativity hypothesis in Proposition 3.12 cannot be dropped.

Corollary 3.15. *Suppose that R is a commutative ring. Then $\langle x_1, \ldots, x_n \rangle = Rx_1 + \cdots + Rx_n$.*

Proof. By Propositions 3.8 & 3.12 $Rx_1 + \cdots + Rx_n$ is an ideal and it certainly contains $x_1, \ldots, x_n$, and hence $\langle x_1, \ldots, x_n \rangle$. On the other hand if I is an ideal containing $x_1, \ldots, x_n$ then $Rx_i \subset I$ and so $Rx_1 + \cdots + Rx_n \subset I + \cdots + I = I$ and the result is proved. □

Remark 3.16. An ideal generated by one element is called a **principal ideal**, and an integral domain in which every ideal is principal is called a **principal ideal domain** or **PID**.

Example 3.17. Example 2.14 and (the second part of) Example 3.6 combine to say that $\mathbb{Z}$ is a PID.

Proposition 3.18. *Suppose that R is a ring and $x \in R$. If $x \in U(R)$ then $\langle x \rangle = R$; if R is commutative and $\langle x \rangle = R$ then $x \in U(R)$.*

Proof. First, since $x \in U(R)$ there is $y \in R$ such that $xy = 1$ and so for all $r \in R$ and any ideal I containing x we have $r = (xy)rx(yr) \in I$, hence $I = R$ and so $\langle x \rangle = R$. On the other hand, if R is commutative and $R = \langle x \rangle$ then $R = xR$ by Proposition 3.12, and there is $y \in R$ such that $xy = 1$. Using commutativity again, $x \in U(R)$. □

Example 3.19. The ideal $\langle 2, X \rangle$ in $\mathbb{Z}[X]$ is the set of polynomials with even constant coefficient. Certainly the polynomials with even constant coefficient form an ideal in $\mathbb{Z}[X]$ containing 2 and X, and conversely every such polynomial is in $\langle 2, X \rangle$ since it can be written in the form $2q + Xp(X)$ for some $p \in \mathbb{Z}[X]$ and constant polynomial $q \in \mathbb{Z}[X]$.

The ideal $\langle 2, X \rangle$ is not principal, so $\mathbb{Z}[X]$ is *not* a PID. To see this, suppose that $p \in \langle 2, X \rangle$ were such that $\langle 2, X \rangle = \langle p \rangle$. By Proposition 3.12 $\langle p \rangle = p(X)\mathbb{Z}[X]$, so there are $q, r \in \mathbb{Z}[X]$ such that $X = p(X)q(X)$ and $2 = p(X)r(X)$, and neither q nor r are the zero polynomial since zero annihilates. Since $\mathbb{Z}$ is an integral domain, Proposition 2.22 (iii) tells us that $0 = \deg 2 = \deg p + \deg r$, so $\deg p = 0$; say $p(X) = a$ for $a \in \mathbb{Z}$. Apply the evaluation homomorphism (Proposition 1.73) $\mathbb{Z}[X] \to \mathbb{Z}$ taking X to 1 to both sides of the first equation to get $1 = aq(1)$. Since $\mathbb{Z}$ is commutative, $a \in U(\mathbb{Z}) = \{-1, 1\}$, and so p has an odd constant coefficient but this contradicts the fact that $p \in \langle 2, X \rangle$.

Kernels

Given a ring homomorphism $\phi : R \to S$, the **kernel of** ϕ is its kernel as a homomorphism of additive groups, that is the set of $r \in R$ such that $\phi(r) = 0_S$.

Remark 3.20. A ring homomorphism $\phi : R \to S$ is injective if and only if $\ker \phi = \{0_R\}$.

Proposition 3.21. *Suppose that $\phi : R \to S$ is a ring homomorphism. Then $\ker \phi$ is an ideal.*

Proof. Since ϕ is a group homomorphism the kernel is an additive subgroup of R. Now suppose $x \in \ker \phi$ and $r \in R$. Then $\phi(xr) = \phi(x)\phi(r) = 0\phi(r) = 0$ since zero annihilates, and similarly $\phi(rx) = 0$. It follows that $xr, rx \in \ker \phi$ so that $\ker \phi$ is an ideal. □

Example 3.22 (Polynomials with common roots as ideals)**.** For R a subring of a commutative ring S and $\lambda \in S$, the set $I := \{p \in R[X] : p(\lambda) = 0\}$ is an ideal since it is the kernel of the evaluation homomorphism $R[X] \to S; p \mapsto p(\lambda)$. We say that λ is a **root** of p if $p \in I$.

Theorem 3.23 (Factor Theorem)**.** *Suppose that R is a commutative ring and $\lambda \in R$. Then $\{p \in R[X] : p(\lambda) = 0\} = \langle X - \lambda \rangle$.*

Proof. Call the left hand set I, and note $X - \lambda \in I$ and I is an ideal, so $\langle X - \lambda \rangle \subset I$ by definition. In the other direction, we require a familiar algebraic identity which follows from the algebra for multiplying polynomials (Proposition 1.12), and the fact that sums of zeros are zero and zeros annihilate:

$$\left(\sum_{i=0}^{n-1} X^i \lambda^{n-1-i}\right)(X - \lambda) = -\lambda^n + \sum_{k=1}^{n-1} \left(\lambda^{n-1-(k-1)} + (-\lambda)\lambda^{n-1-k}\right) X^k + X^n = X^n - \lambda^n$$

for $n \geqslant 1$. Hence if $p \in I$, then by the above, associativity of multiplication, and the fact that right multiplication by $X - \lambda$ is a homomorphism, we have

$$\begin{aligned} p(X) = p(X) - p(\lambda) &= \sum_{n=0}^{d} a_n (X^n - \lambda^n) \\ &= \left(\sum_{n=1}^{d} a_n \left(\sum_{i=0}^{n-1} X^i \lambda^{n-1-i}\right)\right)(X - \lambda) \in \langle X - \lambda \rangle. \end{aligned}$$

The result is proved. □

Proposition 3.24. *Suppose that R is an integral domain and $p \in R[X]^*$ has degree d. Then p has at most d roots in R.*

Proof. We proceed by induction on d. If $d = 0$ then p is a non-zero constant and so has no roots. Now, suppose that $d > 0$ and λ is a root of p. Then there is a polynomial q such that $p(X) = (X - \lambda)q(X)$, and since R is an integral domain Proposition 2.22 (iii) applies so that $\deg q = d - 1$ and so by induction q has at most $d - 1$ roots. Since R is an integral domain, if $\lambda' \in R$ is a root of p then either $\lambda' - \lambda = 0$ or $q(\lambda') = 0$ so that λ' is a root of q. We conclude that p has at most $1 + (d - 1) = d$ roots as claimed. □

Remark 3.25. If R is a non-trivial commutative ring that is not an integral domain then there are elements $a, b \in R^*$ with $ab = 0$. The polynomial $aX \in R[X]$ then has degree 1 but at least two roots: 0 and b.

Quotient rings and the isomorphism theorems

Theorem 3.26 (Quotient Rings)**.** *Suppose that R is a ring and I is an ideal. Then the commutative group R/I may be endowed with a multiplication such that the quotient map $q : R \to R/I; x \mapsto x + I$ is a surjective ring homomorphism with kernel I. If R is commutative then so is this multiplication.*

Proof. I is a subgroup of a commutative group and so normal, and so by the quotient group construction R/I is a commutative group and q is a surjective group homomorphism with kernel I. The key is now to show that $q(xy) = q(x'y')$ whenever $x + I = x' + I$ and $y + I = y' + I$. By distributivity of multiplication and negation we have that $xy - x'y' = (x - x')y + x'(y - y')$. But then $x - x' \in I$ and $y - y' \in I$ and so $xy - x'y' \in Iy + x'I \subset I$ since I is closed under multiplication by *any* element of R (in this case y on the right and x' on the left). We conclude that $q(xy) = q(x'y')$ as required, and so we may define $\widehat{\times}$ on R/I: first, for $u, v \in R/I$ let $x, y \in R$ be such that $q(x) = u$ and $q(y) = v$. Then put $u \widehat{\times} v \coloneqq q(xy)$; this is well-defined by the previous.

For $u, v, w \in R/I$, let $x, y, z \in R$ be such that $u = q(x)$, $v = q(y)$ and $w = q(z)$. Then $(u \widehat{\times} v) \widehat{\times} w = q((xy)z) = q(x(yz)) = u \widehat{\times} (v \widehat{\times} w)$ so that $\widehat{\times}$ is associative. $q(1)q(x) = q(x) = q(x)q(1)$ so $q(1)$ is an identity for $\widehat{\times}$ since q is surjective. Finally, for $q(x) \in R/I$, we have $q(x) \widehat{\times} (q(y) + q(z)) = q(x(y + z)) = q(xy + xz) = q(xy) + q(xz) = q(x) \widehat{\times} q(y) + q(x) \widehat{\times} q(z)$ and since q is surjective it follows that left multiplication by $q(x)$ is a homomorphism. So is right multiplication by a similar argument, and hence (again since q is surjective) it follows that $\widehat{\times}$ distributes over addition.

Finally, we have seen that $q(1)$ is the identity; q is a homomorphism of the additive group by definition of the quotient group; and q is multiplicative by definition. Thus q is a ring homomorphism. Moreover, $\widehat{\times}$ is visibly commutative if the multiplication on R is commutative. The result is proved. □

Remark 3.27. Since the map q above is a surjective ring homomorphism the multiplication on R/I is determined by q: $1_{R/I} = 1 + I$; $(x + I) \times_{R/I} (y + I) = (xy) + I$ for all $x, y \in R$; and if $x \in U(R)$ then $x + I \in U(R/I)$ and $(x + I)^{-1} = x^{-1} + I$, where the first $(\cdot)^{-1}$ is multiplicative inversion in R/I, and the second is in R.

By *the* ring R/I we mean this ring structure.

Example 3.28. For the ring $\mathbb{Z}$ and the ideal $\langle N \rangle = N\mathbb{Z}$ we recover the ring $\mathbb{Z}_N$ from Example 3.1.

Remark 3.29. In the light of this we generalise the notation for modular arithmetic: if R is a ring and I is an ideal in R then we write $x \pmod{I}$ in place of $x + I$ or $q(x)$ (where q is as in Theorem 3.26). The intuition here is that quotient ring R/I is the ring R with the elements of I 'set to zero'.

Example 3.30. Suppose that $\mathbb{F}$ is a field and that I is an ideal in $\mathbb{F}$. Then the map $q : \mathbb{F} \to \mathbb{F}/I$ is a ring homomorphism with kernel I and so by Proposition 1.41 either $\mathbb{F}/I$ is trivial *i.e.* $I = \mathbb{F}$; or this homomorphism is injective and so $I = \{0\}$. It follows that for fields the only two ideals.

⚠ Not every ring with exactly two ideals is a field (see Exercise I.7), but every *commutative* ring with this property is as we shall see later in Proposition 3.46.

Theorem 3.31 (First Isomorphism Theorem)**.** *Suppose that $\phi : R \to S$ is a ring homomorphism. Then $\operatorname{Im}\phi$ is a subring of S; $\ker\phi$ is an ideal in R; and the map*

$$\widetilde{\phi} : R/\ker\phi \to S; x + \ker\phi \mapsto \phi(x)$$

is a well-defined injective ring homomorphism. In particular, $R/\ker\phi \cong \operatorname{Im}\phi$.

Proof. By the First Isomorphism Theorem for groups $\operatorname{Im}\phi$ is an additive subgroup of S; $1_S = \phi(1_R) \in \operatorname{Im}\phi$; and if $x, y \in \operatorname{Im}\phi$ then there are $z, w \in R$ such that $x = \phi(z)$ and $y = \phi(w)$ so $xy = \phi(zw) \in \operatorname{Im}\phi$. The subring test gives the first conclusion. The second conclusion is Proposition 3.21.

The map $\widetilde{\phi}$ is a well-defined injective group homomorphism by the First Isomorphism Theorem for groups. In addition,

$$\begin{aligned}\widetilde{\phi}((x + \ker\phi)(y + \ker\phi)) &= \widetilde{\phi}((xy) + \ker\phi) \\ &= \phi(xy) = \phi(x)\phi(y) = \widetilde{\phi}(x + \ker\phi)\widetilde{\phi}(y + \ker\phi),\end{aligned}$$

and $\widetilde{\phi}(1_R + \ker\phi) = \phi(1_R) = 1_S$. The result is proved. □

Sometimes it is useful not to quotient out by the whole kernel in the First Isomorphism Theorem, for which purpose we have the following lemma.

Lemma 3.32. *Suppose that R is a ring and $I \subset J$ are both ideals of R. Then the map $R/I \to R/J; x + I \mapsto x + J$ is a well-defined ring homomorphism.*

Proof. Call the map π. First, π is well-defined since if $x + I = x' + I$ then $x - x' \in I \subset J$ and so $x + J = x' + J$. $\pi((x + I)(y + I)) = \pi(xy + I) = xy + J = (x + J)(y + J) = \pi(x + I)\pi(y + J)$ and similarly $\pi((x+I)+(y+I)) = \pi((x+y)+I) = (x+y)+J = (x+J)+(y+J) = \pi(x+I)+\pi(y+J)$. Finally, $\pi(1 + I) = 1 + J$ and the result is proved. □

We turn to some consequences of the First Isomorphism Theorem.

Example 3.33. The First Isomorphism Theorem applied to the ring homomorphism $R \to R; x \mapsto x$ gives the isomorphism $R/\{0\} \cong R$.

Example 3.34. For a commutative ring S with a subring R and an element $\lambda \in S$, the image $R[\lambda]$ (defined in Remark 1.74) is a subring by the First Isomorphism Theorem applied to the evaluation homomorphism $R[X] \to S; p \mapsto p(\lambda)$. If $S = R$ then this homomorphism is surjective and we get the isomorphism $R/\langle X - \lambda \rangle \cong R$ by Theorem 3.23.

Example 3.35. In Example 1.35 we saw that $\mathbb{Z}$ had no proper subrings. In the other direction, if R is a ring with no proper subrings then applying the First Isomorphism Theorem to the ring homomorphism $\chi_R : \mathbb{Z} \to R$ provided by Theorem 1.24, we see that the image is a subring of R and so equal to R since R has no proper subrings, and this image is isomorphic to a quotient of $\mathbb{Z}$.

In Example 3.6 we saw every ideal of $\mathbb{Z}$ has the form $\langle N \rangle$ for $N \in \mathbb{N}_0$, so a ring with no proper subrings must be isomorphic to $\mathbb{Z} \cong \mathbb{Z}/\{0\}$ (Example 3.33) or $\mathbb{Z}_N$ for $N \in \mathbb{N}^*$. (A short check confirms that these rings really do not have any proper subrings.)

For a discussion of which (commutative) rings have exactly two proper subrings see [Dob16].

Proposition 3.36. *Suppose that R is an integral domain of characteristic $p \neq 0$. Then p is prime and there is a ring homomorphism $\mathbb{F}_p \to R$ inducing an $\mathbb{F}_p$-vector space structure on the additive group of R in such a way that multiplication on R is bilinear.*

Proof. Let $\chi_R : \mathbb{Z} \to R$ be the homomorphism provided by Theorem 1.24, and suppose that R has characteristic p. If $p = ab$ for $a, b \geqslant 1$ then $0_R = \chi_R(p) = \chi_R(a)\chi_R(b)$, and since R is an integral domain we conclude that $\chi_R(a) = 0$ or $\chi_R(b) = 0$; say the former. Then by definition $a \geqslant p$ and so $a = p$ and $b = 1$. We conclude that p is prime.

The kernel of χ_R contains p and is an ideal in $\mathbb{Z}$. Since $\mathbb{Z}$ is a PID it has the form $\langle N \rangle$ for some $N \in \mathbb{N}_0$, but then $N \mid p$, whence $N = 1$ or $N = p$. If $N = 1$ then $1_R = \chi_R(1) = \chi_R(0) = 0_R$ contradicting the non-triviality of R. We conclude that $N = p$ and the ring $\mathbb{Z}/\langle p \rangle$ is the ring $\mathbb{F}_p$ which is a field (Example 3.7). By the First Isomorphism Theorem there is then an injective ring homomorphism $\mathbb{F}_p \to R$ which induces an $\mathbb{F}_p$-vector space structure on the additive group of R in such a way that right multiplication is $\mathbb{F}$-linear. Since multiplication is commutative, it is $\mathbb{F}$-bilinear. □

Remark 3.37. For a finite field $\mathbb{F}$ the homomorphism $\chi_\mathbb{F} : \mathbb{Z} \to \mathbb{F}$ (from Theorem 1.24) cannot be injective and so the kernel contains a non-zero, and hence positive element so the characteristic is non-zero and hence by the above prime. It follows from this that every finite field has order p^n for some prime p and $n \in \mathbb{N}^*$. In fact it can be shown that there is a field of order p^n for every prime p and $n \in \mathbb{N}^*$, and an accessible proof of this is in [Sou20].

Given an ideal I in R we write $\text{Ideals}_I(R)$ for the set of ideals J in R with $I \subset J$, and $\text{Ideals}(R)(= \text{Ideals}_{\{0\}}(R))$ for the set of ideals of R.

Theorem 3.38 (Relationship between ideals in R and R/I)**.** *Suppose that R is a ring and I is an ideal in R. Then the map*

$$\phi : \text{Ideals}_I(R) \to \text{Ideals}(R/I); I' \mapsto \{x + I : x \in I'\}.$$

is a well-defined inclusion-preserving bijection.

Proof. First, we show the map is well-defined. I' is a subgroup of the additive group of R and the quotient map is a homomorphism of groups so the image of I' under q – which is exactly $\phi(I')$ – is a subgroup of the additive group of R/I. If $x + I \in R/I$ and $y + I \in \phi(I')$ for some $y \in I'$ then $xy, yx \in I'$ and so $(x+I)(y+I) = (xy)+I \in \phi(I')$ and $(y+I)(x+I) = (yx)+I \in \phi(I')$. Thus $\phi(I')$ is genuinely an ideal in R/I.

ϕ is visibly inclusion-preserving. It is an injection since if I' is an ideal containing I and $x \in I'$ then $x + I \subset I'$, hence $\phi(I') = \phi(I'')$ implies $I' = \bigcup \{x + I : x + I \in \phi(I')\} = \bigcup \{x + I : x + I \in \phi(I'')\} = I''$.

Finally, if $J \in \text{Ideals}(R/I)$ then put $I' := \bigcup_{K \in J} K$. $I \subset I'$ since $I = 0_{R/I} \in J$. If $x, y \in I'$ then $x + I, y + I \in J$ and so $(x + (-y)) + I \in J$ (since J is an additive group) and hence $x + (-y) \in I'$. It follows that I' is an additive group by the subgroup test. If $x \in R$ and $y \in I'$ then $(x + I)(y + I) \in J$ and so $(xy) + I \in J$ and $xy \in I'$, and similarly $yx \in I'$ so we see that I' is an ideal. Moreover $\phi(I') = J$, and ϕ is a surjection. □

Remark 3.39. This result also goes by the name of the Correspondence Theorem and sometimes the Fourth Isomorphism Theorem for rings.

Example 3.40. For $N \in \mathbb{N}^*$, $\mathbb{Z}_N$ is a ring in which every ideal is principal. To see this, let $\phi : \text{Ideals}_{N\mathbb{Z}}(\mathbb{Z}) \to \text{Ideals}(\mathbb{Z}_N)$ be the map from the Correspondence Theorem and suppose J is an ideal in $\mathbb{Z}_N$. Since $\mathbb{Z}$ is a PID, $\phi^{-1}(J) = \langle M \rangle$ for some $M \in \mathbb{N}^*$, and furthermore $\langle M \rangle \supset \langle N \rangle$. Since ϕ is a bijection, $J = \phi(\langle M \rangle) = \{Mz + N\mathbb{Z} : z \in \mathbb{Z}\} = \langle M + N\mathbb{Z} \rangle$ is principal.

⚠ $\mathbb{Z}_N$ is *not* a PID unless N is prime since it is not an integral domain.

Proper, prime, and maximal ideals

In Remark 2.15 we arranged non-trivial rings, integral domains, and fields in a hierarchy. This hierarchy is also reflected in ideals. We say that an ideal I in R is **proper** if $I \neq R$, and have the following immediate fact.

Lemma 3.41. *Suppose that R is a commutative ring and I is an ideal in R. Then I is proper if and only if R/I is non-trivial.*

Remark 3.42. We say that an ideal I is **prime** if it is proper and whenever $ab \in I$ we have either $a \in I$ or $b \in I$.

Proposition 3.43. *Suppose that R is a commutative ring and I is an ideal in R. Then I is prime if and only if R/I is an integral domain. In particular R is an integral domain if and only if $\{0_R\}$ is prime.*

Proof. For 'only if' we have $(a+I)(b+I) = 0_{R/I} = I$, so $ab \in I$ and therefore $a \in I$ or $b \in I$ by primality. Consequently $a + I = I = 0_{R/I}$ or $b + I = I = 0_{R/I}$ *i.e.* R/I is an integral domain. (R/I is non-trivial since I is proper.) In the other direction, I is proper since R/I is non-trivial, and if $ab \in I$ then $(a + I)(b + I) = 0_{R/I}$, and $a + I = 0_{R/I} = I$ or $b + I = 0_{R/I} = I$. We conclude $a \in I$ or $b \in I$ as required. □

Example 3.44. The ideal $\langle X \rangle$ is prime in $R[X]$ when R is an integral domain. To see this, suppose that $p(X)q(X) \in \langle X \rangle$. Then by the evaluation homomorphism $p(0)q(0) = 0$, and hence $p(0) = 0$ or $q(0) = 0$ since R is an integral domain. It follows from the Factor Theorem (Theorem 3.23) that $p \in \langle X \rangle$, or $q \in \langle X \rangle$.

Remark 3.45. We say that an ideal I is **maximal** if I is proper and whenever $I \subset J \subset R$ for some ideal J we have $J = I$ or $J = R$.

⚠ Maximal here is maximal with respect to inclusion amongst *proper* ideals; all ideals in R are contained in the ideal R.

Proposition 3.46. *Suppose that R is a commutative ring and I is an ideal in R. Then I is maximal if and only if R/I is a field. In particular R is a field if and only if $\{0\}$ is maximal.*

Proof. Suppose that R/I is a field. Then R/I is non-trivial and so I is proper; suppose J is an ideal with $I \subsetneq J \subset R$. Then there is $x \in J \setminus I$ and since R/I is a field some $y \in R$ such that $xy + I = 1 + I$ whence $1 \in xR + I \subset J$ and so $J = R$, whence I is maximal as claimed.

Conversely, if I is maximal and $x \in R$ has $x + I \neq I$ then $I + xR$ is an ideal properly containing I and so by maximality equals R. It follows that there is some $y \in R$ such that $1 \in xy + I$ whence $(x+I)(y+I) = 1_{R/I}$ so that $U(R/I) = (R/I)^*$ and R/I is a field as required. (R/I is non-trivial as I is proper.) □

Example 3.47. For p a prime number, $\langle p \rangle$ is maximal in $\mathbb{Z}$ since $\mathbb{Z}/\langle p \rangle$ is a field (Example 3.7).

Remark 3.48. Since every field is an integral domain, it follows immediately from this and Proposition 3.43 that every maximal ideal is prime, but this can also be proved directly.

Remark 3.49. ⚠ Although it will turn out that in PIDs all non-zero prime ideals are maximal ideals (essentially Proposition 4.20), this is not true in general *e.g.* $\langle X \rangle$ in $\mathbb{Z}[X]$ is prime, and properly contained in the proper ideal $\langle 2, X \rangle$.

Theorem 3.50 (Krull's Theorem)**.** *Suppose that R is a commutative ring and I is a proper ideal in R. Then there is a maximal ideal J in R containing I.*

Remark 3.51. It is possible to sketch an intuitive iterative argument for the above which can be formalised through a transfinite induction, but the Theorem is more commonly established via Zorn's Lemma following [Zor35]. We shall not give a proof and in fact we could take Theorem 3.50 to be an axiom since it is know to be equivalent to the Axiom of Choice or Zorn's Lemma [Hod79].

4 Divisibility

Principal ideals capture a notion of divisibility: we say that a **divides** b, or a is a **divisor** or **factor** of b, or b is a **multiple** of a, and write $a \mid b$, if $b \in \langle a \rangle$.

Remark 4.1. By definition of generation of ideals, $a \mid b$ if and only if $\langle b \rangle \subset \langle a \rangle$.

Remark 4.2. When R is commutative, Proposition 3.12 tells us that $a \mid b$ if and only if there is $x \in R$ such that $b = xa$.

⚠ If $R = M_2(\mathbb{F})$ and $x \in R$ are as in Example 3.14, then $x \mid 1$ but $1 \neq xr$ for any $r \in R$ for reasons of rank.

Remark 4.3. The structure of ideals means that if $a \mid b_i$ for all $1 \leqslant i \leqslant n$, and $r_1, \ldots, r_n \in R$ then $a \mid b_1 r_1 + \cdots + b_n r_n$.

We say that a and b are **associates** and write $a \sim b$ if $\langle a \rangle = \langle b \rangle$.

Proposition 4.4. *Suppose that R is a commutative ring. Then $\mid$ is reflexive and transitive, and if $x \mid x'$ and $y \mid y'$ then $xy \mid x'y'$. Hence $\sim$ is an equivalence relation, and if $x \sim x'$ and $y \sim y'$ then $xy \sim x'y'$. Furthermore, $0 \mid 0$ if and only if $x = 0$, and $x \mid 1$ if and only if $x \in U(R)$.*

Proof. Reflexivity and transitivity follow immediately from the corresponding facts for subset inclusion. If $x \mid x'$ and $y \mid y'$ then there are elements $a, b \in R$ such that $x' = ax$ and $y' = by$ so $x'y' = (ab)(xy)$, and $xy \mid x'y'$.

Furthermore, $0 \mid 0$, and if $0 \mid x$ then there is $a \in R$ such that $x = 0a = 0$. $x \mid 1$ if and only if $x \in U(R)$ is exactly Proposition 3.18. □

Remark 4.5. ⚠ Ideals depend on the ambient ring and so do $\mid$ and $\sim$ *e.g.* $2 \nmid 3$ in $\mathbb{Z}$, but $2 \mid 3$ in $\mathbb{Q}$.

Remark 4.6. ⚠ Although it is true that if $x = uy$ for a unit $u \in U(R)$ then $x \sim y$, the converse is not in general true even in a commutative ring. (See Exercise II.1.)

Example 4.7. For $r \in R$ and $p(X) = a_0 + a_1 X + \cdots + a_d X^d \in R[X]$ we have $r|p$ in $R[X]$ if and only if $r \mid a_i$ in R for all $0 \leqslant i \leqslant d$ by equating coefficients.

Remark 4.8. We say that c is a **common divisor** of a and b if $c \mid a$ and $c \mid b$, and d is a **greatest common divisor (gcd)** if it is a common divisor, and every common divisor of a and b is a divisor of d. It follows immediately that if d and d' are gcds of a and b then $d \sim d'$.

Similarly we say that m is a **common multiple** of a and b if $a \mid m$ and $b \mid m$, and l is a **least common multiple (lcm)** if it is a common multiple, and every common multiple of a and b is a multiple of l. Again it follows immediately that if l and l' are lcms of a and b then $l \operatorname{dim} l'$.

Remark 4.9. All of this terminology coincides with its usual meaning in $\mathbb{Z}$.

Proposition 4.10. *Suppose that R is commutative ring in which every ideal is principal. Then every pair $a, b \in R$ has a greatest common divisor d and $\langle a \rangle + \langle b \rangle = \langle d \rangle$.*

Proof. Since every ideal in R is principal there is some $d \in R$ such that $\langle a \rangle + \langle b \rangle = \langle a, b \rangle = \langle d \rangle$, and in particular d is a common divisor of a and b. Now if c is a common divisor of a and b, then $a, b \in \langle c \rangle$ and so $\langle d \rangle = \langle a, b \rangle = \langle a \rangle + \langle b \rangle \subset \langle c \rangle + \langle c \rangle = \langle c \rangle$ as required. □

Remark 4.11. We say that an element $x \in R$ is **prime** if $\langle x \rangle$ is a prime ideal; in other notation if $x \neq 1$ and $x \mid ab$ implies $x \mid a$ or $x \mid b$. In particular, if R is an integral domain then Example 3.44 tells us that X is prime in $R[X]$.

Remark 4.12. By induction, given a prime x and a finite list of elements $(y_i)_{i \in I}$ such that $x \mid \prod_{i \in I} y_i$, there is some $i \in I$ such that $x \mid y_i$.

Proposition 4.13. *Suppose that R is an integral domain and $r \in R$ is prime as an element of R. Then r is also prime as an element of $R[X]$.*

Proof. First $r \neq 1$ in R, so $r \notin U(R) = U(R[X])$ and hence $r \neq 1$ in $R[X]$. Suppose that $p(X) = a_0 + a_1 X + \cdots + a_n X^n$ and $q(X) = b_0 + b_1 X + \cdots + b_m X^m$ are such that $r \mid pq$ in $R[X]$ and $r \nmid p$ in $R[X]$ so that there is some minimal $k \in \mathbb{N}_0$ such that $r \nmid a_k$ in R. Suppose that $l \geqslant 0$ and that we have shown $r \mid b_j$ in R for all $j < l$. The coefficient of X^{k+l} in pq is

$$\sum_{j=0}^{k+l} a_j b_{k+l-j} = \sum_{j=0}^{k-1} a_j b_{k+l-j} + a_k b_l + \sum_{j=0}^{l-1} a_{k+l-j} b_j.$$

r divides the left hand side (in R) by hypothesis; it divides the first summand on the right (in R) since $r \mid a_i$ in R for all $0 \leqslant i < k$ by minimality of k; and it divides the last summand (in R) since $r \mid b_j$ in R for all $0 \leqslant j < l$ by the inductive hypothesis. It follows that $r \mid a_k b_l$ in R. But r is prime in R and $r \nmid a_k$ in R by hypothesis, so we conclude $r \mid b_l$ in R. Thus by induction $r \mid b_l$ in R for all $l \in \mathbb{N}_0$ so that $r \mid q$ in $R[X]$ as required. □

Remark 4.14. ⚠ Note that primality is not in general preserved on passage from a subring to a ring: every integral domain is a subring of a field and the only prime in a field is 0.

Remark 4.15. We say that $x \in R$ is **irreducible** if $x \nmid 1$ and whenever $a \mid x$ we have $a \sim x$ or $a \sim 1$. This is equivalent to saying that $\langle x \rangle$ is maximal amongst proper *principal* ideals.

⚠ Irreducible elements can behave in unexpected ways, for example 3 is irreducible in $\mathbb{Z}_6$ but $3^2 \equiv 3 \mod 6$. The next lemma is useful for showing that irreducible elements behave better in integral domains.

Remark 4.16. ⚠ There are many different definitions of irreducibility in the literature, and they are largely equivalent in integral domains. See [AC11] for some examples and discussion.

Lemma 4.17 (Cancellation)**.** *Suppose that R is an integral domain, $w \mid z$ (and z non-zero), and $xz \mid yw$. Then $x \mid y$, and in particular, if $z \sim w$ (are both non-zero) then $xz \sim yw$ if and only if $x \sim y$.*

Proof. Since $w \mid z$ and $xz \mid yw$ there are elements a and b such that $z = aw$ and $bxz = yw$ so $bxaw = yw$ and since w is not a zero-divisor right multiplication by w is injective and so $(ba)x = bxa = y$ and $x \mid y$. □

Proposition 4.18. *Suppose that R is an integral domain and $a, b \in R$. If a and b have a least common multiple l and $ab \neq 0$ then a and b have a greatest common divisor d such that $ab = dl$.*

Proof. Since ab is a common multiple of a and b there is d such that $ab = ld$. Since $a \mid l$, $ad \mid ld = ab$ and so $d \mid b$, similarly $d \mid a$; d is a common divisor. Suppose that $c \mid a$ and $a \mid b$ so $a = gc$ and $b = ch$. Then a and b divide gch, and so $l \mid gch$ and $lc \mid gchc = ab = ld$, whence $c \mid d$. Hence d is a greatest common divisor as claimed. □

Remark 4.19. The assumption that the least common multiple exists is essential. See Exercise II.2 for an example of elements with a gcd but no lcm. It turns out, however, that if *every* pair of elements has a greatest common divisor then every pair of elements has a least common multiple. See, for example, [Cla10, Theorem 40].

Proposition 4.20. *Suppose that R is an integral domain and $x \in R^*$ is prime. Then x is irreducible.*

Proof. Suppose that $x \in R^*$ is prime. First, $x \nmid 1$. Now suppose that $a \mid x$. Then there is $b \in R$ such that $x = ab$. By primality either $x \mid a$ and so $x \sim a$ and we are done; or $x \mid b$ so that $ax \mid ab = x$, and by cancellation $a \mid 1$ since $x \in R^*$, ensuring $a \sim 1$. □

Remark 4.21. Exercise II.2 gives examples to show that even in integral domains, irreducible elements need not be prime.

Remark 4.22. Note that 0_R is always prime in an integral domain R, but it is irreducible if and only if $\langle x \rangle = R$ for all $x \in R^*$, which is true if and only if R is a field.

Proposition 4.23. *Suppose that R is an integral domain such that every pair of elements has a greatest common divisor and $x \in R$ is irreducible. Then x is prime.*

Proof. Suppose $x \mid ab$ and $x \nmid a$. If $b = 0$ then $x \mid b$ as required, so we may suppose $b \in R^*$. By hypothesis xb and ab have a gcd, call it c. Since $b \mid xb$ and $b \mid ab$ we have $b \mid c$, so that $c = db$ for some $d \in R$. Since $db = c \mid xb$ and $db = c \mid ab$, by cancellation we have $d \mid x$ and $d \mid a$. Irreducibility of x tells us that either $d \sim x$ or $d \sim 1$; we cannot have the former since $d \mid a$, but $d \sim x \nmid a$. Hence $d \sim 1$ and so $d \in U(R)$ and $d^{-1}c = b$; in particular, $c \mid b$. But then x is a common factor of xb and ab and so $x \mid c \mid b$ as required. □

Remark 4.24. Usually a positive integer is said to be prime if it is irreducible in the sense of this section. Since $\mathbb{Z}$ is a PID it follows by Propositions 4.10, 4.20 and 4.23 that a positive integer is prime in the usual sense if and only if it is prime in the sense of this section, and there is no conflict in nomenclature.

Primes are particularly important because they ensure a uniqueness of factorisation. To be precise a (possibly empty) vector $(x_1, \dots, x_r)$ is a **factorisation** of an element x if $x \sim x_1 \cdots x_r$; the x_is are called the **factors**, and if all the factors are irreducible then we say that x has a **factorisation into irreducibles**. We say that a factorisation $(x_1, \dots, x_r)$ of x into irreducibles is **unique** if whenever $(y_1, \dots, y_s)$ is factorisation of x into irreducibles there is a bijection $\pi : \{1, \dots, r\} \to \{1, \dots, s\}$ such that $x_i \sim y_{\pi(i)}$ for all $1 \leqslant i \leqslant r$.

Remark 4.25. ⚠ In particular, every unit has a unique factorisation into irreducibles.

Proposition 4.26. *Suppose that R is an integral domain and $x \in R^*$ has a (possibly empty) factorisation in which every factor is prime. Then x has a unique factorisation into irreducibles.*

Proof. Let $(x_1, \dots, x_r)$ be a factorisation of x in which every factor is prime. Since $x \in R^*$, we have $x_1, \dots, x_r \in R^*$, and so by Proposition 4.20 we have that x has a factorisation into irreducibles. We shall prove that if $(y_i)_{i \in I}$ are irreducible elements indexed by a finite set I such that $x \sim \prod_{i \in I} y_i$ then there is a bijection $\pi : \{1, \dots, r\} \to I$ such that $x_i \sim y_{\pi(i)}$ for all $1 \leqslant i \leqslant r$.

We proceed by induction on r. For $r = 0$ we have $\prod_{i \in I} y_i \sim 1$ (by definition of the empty product) and so there is $u \in U(R)$ such that $\prod_{i \in I} y_i = u$. Hence for all $j \in I$, we have $y_j \left(u^{-1} \prod_{i \in I \setminus \{j\}} y_i\right) = 1$ and $y_j \in U(R)$. It follows that I is empty since no unit is irreducible, and we have the base case.

Now, suppose that $r > 0$. Then x_r is prime and $x_r \mid \prod_{i \in I} y_i$ whence there is some $j \in I$ such that $x_r \mid y_j$. But y_j is irreducible and $x_r \not\sim 1$ and so $x_r \sim y_j$. But then $x_1 \cdots x_{r-1} \sim \prod_{i \in I \setminus \{j\}} y_i$ by cancellation, and by the inductive hypothesis there is a bijection $\tilde{\pi} : \{1, \dots, r-1\} \to I \setminus \{j\}$ such that $x_i \sim y_{\tilde{\pi}(i)}$ for all $1 \leqslant i \leqslant r - 1$. Extend this to a bijection $\{1, \dots, r\} \to I$ by setting $\pi(r) = j$ and the result is proved. □

We turn now to the problem of finding factorisations into irreducibles (Proposition 4.23 will then turn these into factorisations in which every factor is prime for use in Proposition 4.26).

We say that a commutative ring R has the **ascending chain condition on principal ideals**[7] or **ACCP** if for every sequence $(d_n)_{n=0}^{\infty}$ of elements of R with $d_{n+1} \mid d_n$ for all $n \in \mathbb{N}_0$, there is some $N \in \mathbb{N}_0$ such that $d_n \sim d_N$ for all $n \geqslant N$. The idea this captures is that we cannot 'keep dividing indefinitely'.

Proposition 4.27. *Suppose that R is a PID. Then R has the ACCP.*

Proof. Suppose that $(d_n)_{n=0}^{\infty}$ has $d_{n+1} \mid d_n$ for all $n \in \mathbb{N}_0$ and let

$$I := \{r \in R : d_n \mid r \text{ for some } n \in \mathbb{N}_0\}.$$

This is an ideal: if $r, s \in I$ then there are $n, m \in \mathbb{N}_0$ such that $d_m \mid r$ and $d_n \mid s$, but $d_{m+n} \mid d_n \mid r$ and $d_{n+m} \mid d_m \mid s$ so $d_{n+m} \mid r - s$ and $r - s \in I$; if $r \in I$ and $s \in R$ then there is $n \in \mathbb{N}_0$ such that $d_n \mid r$ so $d_n \mid rs$ and hence $rs, sr \in I$; and finally $0 \in I$.

Since R is a PID there is some $d \in I$ such that $I = \langle d \rangle$. Since $d \in I$ there is some $N \in \mathbb{N}_0$ such that $d_N \mid d$, but then $d_n \in I$ for all $n \in \mathbb{N}_0$ and so $d_N \mid d \mid d_n$ for all $n \in \mathbb{N}_0$ and hence $d_n \sim d_N$ for all $n \geqslant N$. The result is proved. □

Proposition 4.28. *Suppose that R is an integral domain with the ACCP. Then every $x \in R^*$ has a factorisation into irreducibles.*

Proof. Write $\mathcal{F}$ for the set of elements in R^* that have factorisation into irreducibles so that all units and irreducible elements are in $\mathcal{F}$. $\mathcal{F}$ is closed under multiplication, by design and since R is an integral domain.

Were $\mathcal{F}$ not to be the whole of R^* then there would be some $x_0 \in R^* \setminus \mathcal{F}$. Now create a chain iteratively: at step i suppose we have $x_i \in R^* \setminus \mathcal{F}$. Since x_i is not irreducible and not a unit there is $y_i \mid x_i$ with $y_i \not\sim 1$ and $y_i \not\sim x_i$; let $z_i \in R^*$ be such that $x_i = y_i z_i$. If $z_i \sim x_i$, then $z_i \sim y_i z_i$ and by cancellation $1 \sim y_i$, a contradiction. We conclude $y_i, z_i \not\sim x_i$.

Since $\mathcal{F}$ is closed under multiplication we cannot have both y_i and z_i in $\mathcal{F}$. Let $x_{i+1} \in \{y_i, z_i\}$ such that $x_{i+1} \notin \mathcal{F}$; by design $x_{i+1} \mid x_i$ and $x_{i+1} \not\sim x_i$. This process produces a sequence $\cdots \mid x_2 \mid x_1 \mid x_0$ in which $x_i \not\sim x_{i+1}$ for all $i \in \mathbb{N}_0$ contradicting the ACCP. □

Remark 4.29. Integral domains in which every non-zero element has a factorisation into irreducibles are called **factorisation domains** or **atomic domains**. There are factorisation domains not having the ACCP but these are not easy to construct; the first example was given by Grams in [Gra74].

[7] The reason for the name is that it can also be formulated as saying if $(I_i)_{i \in \mathbb{N}_0}$ is an ascending chain (meaning $I_i \subset I_{i+1}$ for all $i \in \mathbb{N}_0$) of principal ideals then there is some $N \in \mathbb{N}_0$ such that $I_n = I_N$ for all $n \geqslant N$.

Finally, a **unique factorisation domain** or **UFD** is an integral domain in which every $x \in R^*$ has a unique factorisation into irreducibles.

Theorem 4.30. *Suppose that R is a PID. Then R is a UFD.*

Proof. By Propositions 4.27 and 4.28 we have that every $x \in R^*$ has a factorisation into irreducibles. But then every irreducible is prime by Propositions 4.10 and 4.23. The result then follows by Proposition 4.26. □

Remark 4.31. In particular, since $\mathbb{Z}$ is a PID the above gives the Fundamental Theorem of Arithmetic.

Remark 4.32. $\mathbb{Z}[X]$ is an example of a UFD that is not a PID; see Exercise II.7 for details.

Proposition 4.33. *Suppose that R is a UFD and $a \in R^*$. Then there is $s \in \mathbb{N}_0$ such that if $f_1, \ldots, f_r \nmid 1$ and $f_1 \cdots f_r \mid a$ then $r \leqslant s$.*

Proof. Let $p_1 \cdots p_s$ be a factorisation of a into irreducible elements. Since $f_i \nmid 1$ and R is a UFD there is a prime q_i with $q_i \mid f_i$and hence $q_1 \cdots q_r \mid a$. Let $b \in R$ be such that $a = q_1 \cdots q_r b$ and let $q_{r+1}, \ldots, q_t$ be primes such that $q_{r+1} \cdots q_t \sim b$. It follows that $q_1 \cdots q_t \sim a$, and so again since R is a UFD $s = t \geqslant r$. The result is proved. □

The division algorithm and Euclidean domains

A **Euclidean function** on a ring R is a function $f : R^* \to \mathbb{N}_0$ such that

- $f(a) \leqslant f(b)$ whenever $a \mid b$ (both non-zero);
- and if $a, b \in R^*$ then either $b \mid a$, or there are $q \in R$, $r \in R^*$ such that $a = bq + r$ and $f(r) < f(b)$.

We say that an integral domain R is a **Euclidean domain** if R supports at least one Euclidean function.

Remark 4.34. Keating [Kea98, p17] uses an even stronger definition of Euclidean function f requiring that $f(ab) = f(a)f(b)$ whenever $a, b \in R^*$. This is a genuinely stronger definition, meaning there are Euclidean domains in our sense but not in the sense of Keating, though this is a recent discovery: [CNT19, Theorem 1.3].

Example 4.35 (Fields are Euclidean domains)**.** Suppose that $\mathbb{F}$ is a field and let $f : \mathbb{F}^* \to \mathbb{N}_0$ be the constant function 1. Since $f(a) = f(b)$ for all a and b, and every two non-zero units divide each other in a field, f is a Euclidean function for $\mathbb{F}$ and so $\mathbb{F}$ is a Euclidean domain.

Example 4.36 ($\mathbb{Z}$ is a Euclidean domain)**.** If $a, b \in \mathbb{Z}^*$ and $b \nmid a$ then let bq be (one of) the multiple(s) of b nearest to a. Then $r := a - bq$ has $|r| < |b|$, and $|\cdot|$ is a Euclidean function on $\mathbb{Z}$ and $\mathbb{Z}$ is a Euclidean domain. (It certainly has $|a| \leqslant |b|$ whenever $a \mid b$.)

⚠ Note that there were *two* choices for bq and hence for r in the case that $b \nmid a$.

Example 4.37 (Polynomial rings over fields are Euclidean domains)**.** Suppose that $\mathbb{F}$ is a field and $a, b \in \mathbb{F}[X]^*$. Let $\mathcal{P} := \{a + bq : q \in \mathbb{F}[X]\}$, and note that if $b \nmid a$ then $\mathcal{P}$ does not include the zero polynomial.

If $b \nmid a$, we let $r \in \mathcal{P}$ be a polynomial of minimal degree. If $\deg r \geqslant \deg b$, then we may let λ be the ratio of the lead coefficient of r to that of b and note that $r(X) - \lambda X^{\deg r - \deg b} b(X) \in \mathcal{P}$ and has strictly smaller degree than r, a contradiction. It follows that $\deg r < \deg b$. Finally $\deg p \leqslant \deg q$ whenever $p \mid q$ by Proposition 2.22 (iii), and so deg is a Euclidean function and $\mathbb{F}[X]$ is a Euclidean domain.

Remark 4.38. Suppose that f is a Euclidean function on an integral domain R such that $f(a) \leqslant f(ab)$ for all $a, b \in R^*$, and for all $a, b \in R^*$ either $b \mid a$ or there is a unique pair $(q, r) \in R \times R^*$ with $a = bq + r$ and $f(r) < f(b)$. Then either R is itself a field or $R = \mathbb{F}[X]$ for a field $\mathbb{F}$. A short proof of this may be found in [Jod67]. In particular, since $\mathbb{Z}$ is neither a field nor a polynomial ring over a field this explains why we had to make a choice in Example 4.36.

Proposition 4.39. *Suppose that R is a Euclidean domain. Then R is a PID.*

Proof. Let f be a Euclidean function on R and suppose I is a non-zero ideal. Let $x \in I$ have $f(x)$ minimal, and suppose that $y \in I$. If $y \notin \langle x \rangle$ then there is $q \in R$ and $r \in R^*$ with $y = qx + r$ and $f(r) < f(x)$ so that $r \in I$, contradicting minimality of $f(x)$. □

Remark 4.40. In particular if $\mathbb{F}$ is a field then the ring $\mathbb{F}[X]$ is a PID.

Remark 4.41. There are examples of PIDs which are not Euclidean domains, for example $\mathbb{Z}[\theta]$ where $\theta^2 - \theta + 5 = 0$ (a proof may be found in [Con, Theorem 5.13]), and $\mathbb{R}[X, Y]/\langle X^2 + Y^2 + 1 \rangle$ (a proof may be found in [Con, Theorem 5.14]).

5 Fields and adjoining elements

A field $\mathbb{K}$ is an **extension field** of a field $\mathbb{F}$ if there is a ring homomorphism $\phi : \mathbb{F} \to \mathbb{K}$. Formally, ϕ is a **field extension** written $\mathbb{K} : \mathbb{F}$.

Remark 5.1. ⚠ The notation $\mathbb{K} : \mathbb{F}$ refers only to the domain and codomain of ϕ. The actual map will often just be inclusion and indeed by relabelling the elements of $\mathbb{F}$ we can always assume that $\mathbb{F}$ is a subfield of $\mathbb{K}$ because ring homomorphisms between fields are injective (Proposition 1.41).

Remark 5.2. The map ϕ induces an $\mathbb{F}$-vector space structure on $\mathbb{K}$ such that multiplication in $\mathbb{K}$ is bilinear. We call the $\mathbb{F}$-dimension of $\mathbb{K}$ w.r.t. this vector space structure the **degree** of the field extension and denote it $|\mathbb{K} : \mathbb{F}|$.

Theorem 5.3 (Tower Law)**.** *Suppose that $\phi : \mathbb{K} \to \mathbb{L}$ and $\psi : \mathbb{F} \to \mathbb{K}$ are field extensions. Then $\phi \circ \psi : \mathbb{F} \to \mathbb{L}$ is a field extension and if either $|\mathbb{L} : \mathbb{F}| < \infty$ or $|\mathbb{L} : \mathbb{K}|, |\mathbb{K} : \mathbb{F}| < \infty$ then $|\mathbb{L} : \mathbb{F}| = |\mathbb{L} : \mathbb{K}||\mathbb{K} : \mathbb{F}|$.*

Proof. First, the composition of homomorphisms is a homomorphism so that $\phi \circ \psi$ is a field extension. Since all ring homomorphisms between fields are injective (Proposition 1.41), by relabelling we may assume that $\mathbb{F}$ is a subfield of $\mathbb{K}$ and $\mathbb{K}$ is a subfield of $\mathbb{L}$. We do this to make the notation simpler.

Let $e_1, \ldots, e_n$ be a basis for $\mathbb{L}$ as a vector space over $\mathbb{K}$, and let $f_1, \ldots, f_m$ be a basis for $\mathbb{K}$ as a vector space over $\mathbb{F}$. Now, for $x \in \mathbb{L}$ there are scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ such that $x = \lambda_1 e_1 + \cdots + \lambda_n e_n$, and since $f_1, \ldots, f_m$ is spanning, for each $1 \leqslant j \leqslant n$ there are scalars $\mu_{1,j}, \ldots, \mu_{m,j} \in \mathbb{F}$ such that $\lambda_j = \mu_{1,j} f_1 + \cdots + \mu_{m,j} f_m$. Hence $x = \sum_{j=1}^{n} \sum_{i=1}^{m} \mu_{i,j} f_i e_j$, so by have that $(f_i e_j)_{i=1,j=1}^{m,n}$ is an $\mathbb{F}$-spanning subset of $\mathbb{K}$. Now suppose $\mu_{1,1}, \ldots, \mu_{m,n} \in \mathbb{F}$ are such that $\sum_{j=1}^{n} \sum_{i=1}^{m} \mu_{i,j} f_i e_j = 0_{\mathbb{L}}$. Then $\sum_{j=1}^{n} \left(\sum_{i=1}^{m} \mu_{i,j} f_i\right) e_j = 0_{\mathbb{L}}$, but $\sum_{i=1}^{m} \mu_{i,j} f_i \in \mathbb{K}$ for each $1 \leqslant j \leqslant n$ and since $e_1, \ldots, e_n$ are $\mathbb{K}$-linearly independent we have $\sum_{i=1}^{m} \mu_{i,j} f_i = 0_{\mathbb{K}}$ for all $1 \leqslant j \leqslant n$. But now $f_1, \ldots, f_m$ are $\mathbb{F}$-linearly independent and so $\mu_{i,j} = 0_{\mathbb{F}}$ for all $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$. It follows that $(f_i e_j)_{i=1,j=1}^{m,n}$ is a basis for $\mathbb{L}$ as an $\mathbb{F}$-vector space and the result follows. □

Remark 5.4. If $\mathbb{F}$ is a finite field, and $|\mathbb{K} : \mathbb{F}| = n$, $|Ł : \mathbb{K}| = m$, and $|Ł : \mathbb{F}| = k$ then $|\mathbb{K}| = |\mathbb{F}|^n$, $|Ł| = |\mathbb{K}|^m$, and $|Ł| = |\mathbb{F}|^k$ from which it follows that $k = nm$. The Tower Law extends this to infinite fields.

Example 5.5. Suppose that $\mathbb{L}$ is a field of order 8, and $\mathbb{K}$ is a field of order 4. We can use the Tower Law to show that there is no ring homomorphism $\phi : \mathbb{K} \to \mathbb{L}$ – in words $\mathbb{K}$ is not isomorphic to a subfield of $\mathbb{L}$. First ,since $\mathbb{K}$ has order 4, we know from Proposition 3.36 that there is a ring homomorphism $\pi : \mathbb{F}_2 \to \mathbb{K}$ hence $\mathbb{K} : \mathbb{F}_2$ is a degree 2 field extension, and then $\phi \circ \pi : \mathbb{F}_2 \to \mathbb{L}$ is a degree 3 field extension. By the Tower Law we have $|\mathbb{F}_8 : \mathbb{F}_4| \times 2 = |\mathbb{F}_8 : \mathbb{F}_4||\mathbb{F}_4 : \mathbb{F}_2| = |\mathbb{F}_8 : \mathbb{F}_2| = 3$ which leads to a contradiction.

We shall see later (in Examples 5.12 & 5.13) that there actually *are* fields of order 4 and 8 respectively (a special case of Remark 3.37), and in fact they are unique up to ring isomorphism.

Theorem 5.6. *Suppose that $\mathbb{F}$ is a field and $f \in \mathbb{F}[X]$ is irreducible of degree d. Then $\mathbb{K} := \mathbb{F}[X]/\langle f \rangle$ is an extension field of $\mathbb{F}$ by the map $\mathbb{F} \to \mathbb{K}; \lambda \mapsto \lambda + \langle f \rangle$, and writing $\alpha := X + \langle f \rangle$, $1_{\mathbb{K}}, \alpha, \ldots, \alpha^{d-1}$ is a basis for $\mathbb{K}$ in this $\mathbb{F}$-vector space structure.*

Proof. $\mathbb{F}[X]$ is a PID (Remark 4.40) and hence the fact that $\langle f \rangle$ is maximal amongst proper principal ideals means it is maximal amongst *all* proper ideals and Proposition 3.46 tells us that $\mathbb{K} = \mathbb{F}[X]/\langle f \rangle$ is a field. The given map is formed by composing the inclusion map $\mathbb{F} \to \mathbb{F}[X]$ and the quotient map $\mathbb{F}[X] \to \mathbb{F}[X]/\langle f \rangle$ and so is a ring homomorphism, and hence $\mathbb{K}$ an extension field of $\mathbb{F}$ by the given map.

The elements $1_{\mathbb{K}}, \alpha, \ldots, \alpha^{d-1}$ are $\mathbb{F}$-independent in $\mathbb{K}$: indeed, suppose that $a_0, \ldots, a_{d-1} \in \mathbb{F}$ have $0_{\mathbb{K}} = a_0.1_{\mathbb{K}} + a_1.\alpha + \cdots + a_{d-1}.\alpha^{d-1}$. This says exactly that $f \mid a_0 + a_1X + \cdots + a_{d-1}X^{d-1}$. If the right hand side is non-zero then it has degree strictly smaller than d; a contradiction. Hence the right is $0_{\mathbb{F}[X]}$ and so $a_0, \ldots, a_{d-1} = 0_{\mathbb{F}}$ as required.

On the other hand, if $f(X) = a_0 + a_1X + \cdots + a_dX^d$ then every $\beta \in \mathbb{K}$ has a polynomial $p(X) = b_0 + b_1X + \cdots + b_nX^n \in \mathbb{F}[X]$ such that $\beta = p(X) + \langle f \rangle$. By the division algorithm for $\mathbb{F}[X]$ (Example 4.37), either $p \in \langle f \rangle$ (and so $\beta = 0_{\mathbb{K}}$) or there is some $q \in \mathbb{F}[X]$ and $r \in \mathbb{F}[X]^*$ with $\deg r < \deg f = d$ such that $p(X) = q(X)f(X) + r(X)$. Then $\beta = r(X) + \langle f \rangle$, and writing $r(X) = c_0 + c_1X + \cdots + c_{d-1}X^{d-1}$ for $c_0, \ldots, c_{d-1} \in \mathbb{F}$ we have $\beta = c_0.1_{\mathbb{K}} + c_1.\alpha + \cdots + c_{d-1}.\alpha^{d-1}$, and hence $1_{\mathbb{K}}, \alpha, \ldots, \alpha^{d-1}$ is a spanning set.

It follows that $1_{\mathbb{K}}, \alpha, \ldots, \alpha^{d-1}$ is a basis and the result is proved. □

In view of Theorem 5.6 it becomes important to identify irreducible polynomials in $\mathbb{F}[X]$.

Proposition 5.7. *Suppose that $\mathbb{F}$ is a field and $f \in \mathbb{F}[X]$. Then $\deg f = 1$ if and only if f is irreducible and has a root in $\mathbb{F}$.*

Proof. Suppose $\deg f = 1$. Then $f(X) = aX + b$ for $a \in \mathbb{F}^*$ and so f has $-b/a$ as a root. Since $\deg f \neq 0$, $f \not\sim 1$ and if $g \mid f$ then $\deg g \leqslant \deg f = 1$ so either $\deg g = 0$ and $g \sim 1$; or $\deg g = 1$ and writing $f = hg$ for some $h \in \mathbb{F}[X]^*$ we have $\deg h = 0$ by Proposition 2.22 (iii), and so $h(X) = a \in U(\mathbb{F})$ and in particular $g \sim f$.

Suppose that f is irreducible and has a root $\lambda \in \mathbb{F}$. By the Factor Theorem $X - \lambda \mid f$, and by irreducibility $X - \lambda \sim f$ whence $\deg f = 1$. □

Example 5.8. The irreducible polynomials in $\mathbb{C}[X]$ are exactly the degree 1 polynomials. The Fundamental Theorem of Algebra tells us that every non-constant polynomial in $\mathbb{C}[X]$ has a root in $\mathbb{C}$, and the result follows by Proposition 5.7.

Lemma 5.9. *Suppose that $\mathbb{F}$ is a field and $f \in \mathbb{F}[X]$ a non-constant polynomial with degree at most 3 and no root in $\mathbb{F}$. Then f is irreducible.*

Proof. Suppose f is *not* irreducible. Since f is non-constant, $f \not\sim 1$, and so for f not to be irreducible there must be some g with $g \mid f$, $g \not\sim 1$ and $g \not\sim f$. Write $f = gh$ for some $h \in \mathbb{F}[X]^*$. Since $g \not\sim 1$ we have that $\deg g \neq 0$ and since $g \not\sim f$ we have $\deg h \neq 0$. Since $\deg g + \deg h \leqslant 3$ it follows that $\deg g = 1$ or $\deg h = 1$; in the former case g has a root by Proposition 5.7, and in the latter h does. In either case $f = gh$ has a root leading to a contradiction. □

Example 5.10. The polynomial $(X^2+1)^2$ in $\mathbb{R}[X]$ is a non-constant polynomial with degree 4 and no root in $\mathbb{R}$ which is *not* irreducible so the 3 in Lemma 5.9may not be increased.

Example 5.11. The evaluation polynomial $\phi : \mathbb{R}[X] \to \mathbb{C}; p \mapsto p(i)$ is a surjective ring homomorphism so by the First Isomorphism Theorem $\mathbb{C} \cong \mathbb{R}[X]/\ker\phi$. Since $\mathbb{R}[X]$ is a PID, $\ker\phi = \langle p\rangle$ for some $p \in \mathbb{R}[X]$. Since $i^2+1=0$ we have $p \mid X^2+1$. By Lemma 5.9 X^2+1 is irreducible over $\mathbb{R}$, and hence either $p \sim 1$ in which case $\ker\phi = \mathbb{R}[X]$, but $\mathbb{C}$ is not-trivial so this is a contradiction; or else $p \sim X^2+1$. It follows that $\mathbb{C} \cong \mathbb{R}[X]/\langle X^2+1\rangle$.

Alternatively, we may construct $\mathbb{C}$ by defining it to be the ring $\mathbb{R}[X]/\langle X^2+1\rangle$ and using Theorem 5.6 to establish its properties as in Example 1.43.

Example 5.12. The polynomial X^2+X+1 is irreducible in $\mathbb{F}_2[X]$ by Lemma 5.9 since neither 0 nor 1 are roots. On the other hand there are only four degree 2 polynomials in $\mathbb{F}_2[X]$, and the other three are X^2, X^2+X and X^2+1 which visibly have roots of 0, 0 (and 1), and 1 respectively. Hence none of these is irreducible.

By Theorem 5.6, the ring $\mathbb{F}_2[X]/\langle X^2+X+1\rangle$ is then a field of order 4 which is denoted $\mathbb{F}_4$. ⚠ This field is *not* equal to the ring $\mathbb{Z}_4$ – indeed the latter is not even an integral domain since $2^2 \equiv 0 \pmod 4$ but $2 \not\equiv 0 \pmod 4$.

Example 5.13. The polynomial X^3+X+1 is irreducible in $\mathbb{F}_2[X]$ by Lemma 5.9 since neither 0 nor 1 are roots. By Theorem 5.6, the ring $\mathbb{F}_2[X]/\langle X^3+X+1\rangle$ is then a field of order 8 which is denoted $\mathbb{F}_8$.

Finding irreducible polynomials is little like finding primes in the integers, and there are various tests for irreducibility which can help in this endeavour.

We say that $f \in \mathbb{Z}[X]$ is **primitive** if there is no prime dividing all the coefficients of f.

Remark 5.14. If f is primitive and of degree 0 then f is a unit in $\mathbb{Z}[X]$ since $\mathbb{Z}$ is a UFD (and so every non-unit has a prime factor).

Theorem 5.15 (Gauss' Lemma)**.** *Suppose that $f \in \mathbb{Z}[X]$. Then f is non-constant and irreducible in $\mathbb{Z}[X]$ if and only if f is primitive and irreducible in $\mathbb{Q}[X]$.*

Proof. Suppose that f is irreducible in $\mathbb{Z}[X]$. This immediately tells us that f is primitive since if p were a prime dividing all the coefficients of f then $p \mid f$ in $\mathbb{Z}[X]$. Since $p \nmid 1$ we conclude that $p \sim f$ (in $\mathbb{Z}[X]$) by irreducibility of f, contradicting the fact that f is non-constant.

Now, suppose that $f = gh$ for $g, h \in \mathbb{Q}[X]$. Then let $\lambda \in \mathbb{N}^*$ be minimal such that there is $q \in \mathbb{Q}^*$ with $\lambda q^{-1} g$ and qh both in $\mathbb{Z}[X]$. Suppose that $p \in \mathbb{Z}$ is prime with $p \mid \lambda$. Then p is prime as a constant polynomial in $\mathbb{Z}[X]$ and since $p \mid \lambda f = (\lambda q^{-1} g)(qh)$, we have $p \mid \lambda q^{-1} g$ or $p \mid qh$ (both in $\mathbb{Z}[X]$). The former contradicts minimality of λ directly, and the latter once we note that $(q/p)h \in \mathbb{Z}[X]$ and $(\lambda/p)(q/p)^{-1} g = \lambda q^{-1} g \in \mathbb{Z}[X]$. We conclude that λ

has no prime factors and hence (since $\mathbb{Z}$ is a UFD) is a unit. Thus $q^{-1}g \mid f$ in $\mathbb{Z}[X]$ and so by irreducibility of f in $\mathbb{Z}[X]$ we conclude that either $q^{-1}g \sim 1$ or $q^{-1}g \sim f$ in $\mathbb{Z}[X]$. Hence either $g \sim 1$ in $\mathbb{Q}[X]$ or $g \sim f$ in $\mathbb{Q}[X]$ and finally, since f is non-constant we have $f \not\sim 1$ in $\mathbb{Q}[X]$ and so f is irreducible in $\mathbb{Q}[X]$.

Conversely, suppose $f \in \mathbb{Z}[X]$ is primitive and irreducible in $\mathbb{Q}[X]$. First, $f \not\sim 1$ in $\mathbb{Q}[X]$ and so f is non-constant. Suppose $g \mid f$ in $\mathbb{Z}[X]$. By irreducibility of f in $\mathbb{Q}[X]$, either $g \sim 1$ in $\mathbb{Q}[X]$ so $\deg g = 0$, and since f is primitive $g \sim 1$ in $\mathbb{Z}[X]$; or $g \sim f$ in $\mathbb{Q}[X]$, then $\deg g = \deg f$ and writing $f = gh$ for $h \in \mathbb{Z}[X]$ we have $\deg h = 0$, and since f is primitive $h \sim 1$ in $\mathbb{Z}[X]$, whence $g \sim f$ in $\mathbb{Z}[X]$. The result is proved. □

Proposition 5.16 (Eisenstein's Criterion)**.** *Suppose that $f(X) = a_n X^n + \cdots + a_1 X + a_0$ is a primitive polynomial in $\mathbb{Z}[X]$ and p is a prime such that $p \mid a_i$ for all $0 \leqslant i < n$; $p \nmid a_n$; and $p^2 \nmid a_0$. Then f is irreducible in $\mathbb{Z}[X]$.*

Proof. Suppose that $f = gh$ for $g, h \in \mathbb{Z}[X]$. The quotient map $\mathbb{Z} \to \mathbb{F}_p$ is a homomorphism so there is an evaluation homomorphism $\phi : \mathbb{Z}[X] \to \mathbb{F}_p[X]$ taking X to X. In particular, note that

$$\phi(f) = \phi(g)\phi(h) \text{ and } \deg q \geqslant \deg \phi(q) \text{ whenever } \phi(q) \in \mathbb{F}_p[X]^*.$$

Since $p \mid a_i$ for all $i < n$ and $p \nmid a_n$ we have $\phi(f) \sim X^n$.

Since $X \in \mathbb{F}_p[X]$ is prime it follows that $\phi(g) \sim X^i$ and $\phi(h) \sim X^{n-i}$ (either by induction, or because $\mathbb{F}_p[X]$ is a UFD). If $i > 0$ then $\phi(g)$ has zero constant coefficient and so p divides the constant coefficient of g. a_0 is the product of the constant coefficients of g and h and since $p^2 \nmid a_0$ we conclude that p does not divide the constant coefficient of h *i.e.* $i = n$. But then $\deg g \geqslant \deg \phi(g) = n$, and $n = \deg f = \deg g + \deg h$, so $\deg h = 0$. Since f is primitive, h is then a unit and so $g \sim f$. The case $i = 0$ is handled similarly and has $g \sim 1$ □

Example 5.17. For $n \in \mathbb{N}^*$, the polynomial $X^n - 2$ is irreducible in $\mathbb{Z}[X]$ by Eisenstein's Criterion with the prime 2 since it is visibly primitive (with the lead coefficient being 1). It is non-constant and so by Gauss' Lemma is irreducible in $\mathbb{Q}[X]$. By Theorem 5.6, $\mathbb{Q}[X]/\langle X^n - 2\rangle : \mathbb{Q}$ is a degree n field extension.

6 Modules

Modules can be viewed in a variety of ways. First, we shall think of them as vector spaces with the field replaced by a ring. Concretely, suppose that R is a ring. An **R-module** is a commutative group called the **additive group**, with operation $+$ and identity 0, and a map $.: R \times M \to M; (r, x) \mapsto r.x$ called the **scalar multiplication**, such that

(M1) $1.x = x$ for all $x \in M$;

(M2) $r.(s.x) = (rs).x$ for all $r, s \in R$ and $x \in M$;

(M3) $(r + s).x = (r.x) + (s.x)$ for all $r, s \in R$ and $x \in M$;

(M4) $r.(x + y) = (r.x) + (r.y)$ for all $r \in R$ and $x, y \in M$.

The elements of the additive group are called **vectors**, the elements of the ring are called **scalars**, and M is sometimes said to be a **module over** R. We shall disambiguate between multiple modules with subscripts writing, for example, $+_M$ or 0_M instead of $+$ and 0 above.

The operation $+$ is the **addition** of the module and 0 is the **zero** of the module. For each $x \in M$ we write $-x$ for the unique inverse of x, and the map $M \to M; x \mapsto -x$ is the **negation** of the module.

Remark 6.1. Identities are self-inverse so $-0 = 0$; double inversion is the identity map so $-(-x) = x$ for all $x \in M$; and negation is a homomorphism. Since addition is an associative and commutative binary operation with identity, for a finite set I, and elements $x_i \in M$ for all $i \in I$, we can give a definition of $\sum_{i \in I} x_i$ such that if I is empty the sum is 0_M; if I is a singleton the sum is x_i; and for any partition $\mathcal{P}$ of I we have

$$\sum_{i \in I} x_i = \sum_{P \in \mathcal{P}} \left(\sum_{i \in P} x_i \right).$$

We use the terminology 'change of variables' in the same way as for rings. (*c.f.* Remarks 1.3, 1.4, 1.5, 1.6, & 1.7.)

Remark 6.2. ⚠ If the scalar multiplication is clear we simply speak of the R-module M.

Remark 6.3. (M4) says exactly that for $r \in R$ the map $M \to M; x \mapsto r.x$ is a group homomorphism of the additive group of M, so $r.0_M = 0_M$ and $r.(-x) = -(r.x)$ for all $x \in M$. (M3) says exactly that for $x \in M$ the map $R \to M; r \mapsto r.x$ is a group homomorphism from the additive group of R to the additive group of M, so $0_R.x = 0_M$ and $(-r).x = -(r.x)$ for all $r \in R$. (*c.f.* Remark 1.8.)

Example 6.4 (Vector spaces as modules)**.** Given a field $\mathbb{F}$, a vector space V is exactly an $\mathbb{F}$-module, with the two notions of scalar multiplication coinciding.

⚠ In a vector space V, for $v \neq 0_V$ the map $\mathbb{F} \to V; \lambda \mapsto \lambda.v$ is injective, but in more general modules it need not be.

Example 6.5 (Zero module)**.** For any ring R, the trivial group – usually denoted $\{0\}$ in this context – and the scalar multiplication defined by $r.0 := 0$ for all $r \in R$ is a module called the **zero (R-)module**.

Proposition 6.6. *Suppose that R is a ring. Then the ring multiplication of R is a scalar multiplication of the ring R on the additive group of R.*

Proof. (M1) follows since 1_R is an identity of ring multiplication; (M2) by associativity of ring multiplication; (M3) since all right multiplication maps on a ring are homomorphisms of the additive group; (M4) since all left multiplication maps on a ring are homomorphisms. □

Remark 6.7. We call the R-module structure of the above proposition *the* R**-module** R. The additive group of R as a ring and of R as an R-module are the same.

⚠ In general there can be multiple R-module structures on R. Indeed, in Example 1.44 we noted that complex conjugation is a ring homomorphism from $\mathbb{C}$ to itself and this induces a $\mathbb{C}$-vector space structure on the additive group of $\mathbb{C}$. A $\mathbb{C}$-vector space structure is exactly a $\mathbb{C}$-module structure (Example 6.4), and in this particular structure $i.1 = \overline{i} = -i$, whereas in the $\mathbb{C}$-module $\mathbb{C}$ (using the definition at the start of the remark) we have $i.1 = i$.

Write $\mathrm{Scalar}_R(M)$ for the set of functions $. : R \times M \to M$ satisfying the axioms (M1)–(M4); and $\mathrm{RingHom}(R, S)$ the set of ring homomorphisms $R \to S$. ⚠ Neither of these pieces of notation is standard.

Theorem 6.8. *Suppose that R is a ring and M is a commutative group. Then*

$$\mathcal{C} : \mathrm{Scalar}_R(M) \to \mathrm{RingHom}(R, \mathrm{Hom}(M, M))$$

$$. : R \times M \to M \mapsto \mathcal{C}(.) : \begin{array}{ccc} R & \to & \mathrm{Hom}(M, M) \\ r & \mapsto & (M \to M; x \mapsto r.x) \end{array}$$

and

$$\mathcal{U} : \mathrm{RingHom}(R, \mathrm{Hom}(M, M)) \to \mathrm{Scalar}_R(M)$$

$$\phi : R \to \mathrm{Hom}(M, M) \mapsto \mathcal{U}(\phi) : \begin{array}{ccc} R \times M & \to & M \\ (r, x) & \mapsto & \phi(r)(x) \end{array}$$

are well-defined and inverses of each other.

Proof. Suppose that $. : R \times M \to M$ is a scalar multiplication. Then for $r \in R$ the map $M \to M; x \mapsto r.x$ is a group homomorphism by (M4), so $\mathcal{C}(.)$ maps into $\mathrm{Hom}(M, M)$. The fact that $\mathcal{C}(.)$ is a ring homomorphism then follows from (M1)–(M3).

In the other direction, if $\phi : R \to \mathrm{Hom}(M, M)$ is a ring homomorphism then the map $\mathcal{U}(\phi) : R \times M \to M; (r, x) \mapsto \phi(r)(x)$ satisfies (M4) since $\phi(r)$ is a homomorphism of M, and (M1)–(M3) since ϕ is a ring homomorphism. In other words $\mathcal{U}(\phi)$ is a scalar multiplication.

Finally, $r\mathcal{U}(\mathcal{C}(.))x = \mathcal{C}(.)(r)(x) = r.x$ and $\mathcal{C}(\mathcal{U}(\phi))(r)(x) = \phi(r)(x)$ for all $r \in R$ and $x \in M$ and so the maps $\mathcal{C}$ and $\mathcal{U}$ are inverses of each other. □

Remark 6.9. The map $\mathcal{C}$ is called **currying** and the map $\mathcal{U}$ is called **uncurrying**.

Remark 6.10. Currying the scalar multiplication of the R-module R gives the homomorphism of Theorem 1.52.

Example 6.11 (Commutative groups as modules)**.** Suppose that M is a commutative group. Then by Theorem 1.24 there is a homomorphism $\mathbb{Z} \to \mathrm{Hom}(M, M)$ which by uncurrying endows M with the structure of a $\mathbb{Z}$-module. The scalar multiplication here is familiar: $0.x = 0$ (as always), and for $n \in \mathbb{N}^*$, $n.x$ is the n-fold sum of x with itself, and $(-n).x$ is the n-fold sum of $-x$ with itself.

The fact that the homomorphism of Theorem 1.24 is unique and the currying and uncurrying maps are inverses of each other means that there is a *unique* way to make the group M into a $\mathbb{Z}$-module.

Example 6.12 (Polynomial rings as modules)**.** Suppose that R is a commutative ring. The scalar multiplication from the $R[X]$-module $R[X]$ (as defined in Proposition 6.6) curries to give a homomorphism $R[X] \to \mathrm{Hom}(R[X], R[X])$, and since R is a subring of $R[X]$ this composes with the inclusion map $R \to R[X]$ to give a ring homomorphism $R \to \mathrm{Hom}(R[X], R[X])$. This uncurries to a scalar multiplication of R on $R[X]$ with $r.(a_0 + \cdots + a_n X^n) = (ra_0) + \cdots + (ra_n)X^n$.

Example 6.13 (Vector spaces with an endomorphism as modules)**.** Suppose that V is an $\mathbb{F}$-vector space and $T : V \to V$ is $\mathbb{F}$-linear (this is the eponymous endomorphism). Then by currying the vector space structure gives a homomorphism $\phi : \mathbb{F} \to \mathrm{Hom}(V, V)$. Since T is a group homomorphism of the additive group of V we have $T \in \mathrm{Hom}(V, V)$, and since $T(\lambda.v) = \lambda.T(v)$ for all $\lambda \in \mathbb{F}$ and $v \in V$ we have that T commutes with the image of ϕ and so by Proposition 1.73 there is an evaluation homomorphism $\mathbb{F}[X] \to \mathrm{Hom}(V, V)$ taking X to T. By uncurrying this gives V the structure of an $\mathbb{F}[X]$-module. Concretely the scalar multiplication has

$$(a_0 + a_1 X + \cdots + a_d X^d).v = a_0 v + a_1 Tv + \cdots a_d T^d v \text{ for all } p \in \mathbb{F}[X] \text{ and } v \in V.$$

The uniqueness of the homomorphism in Proposition 1.73 ensures that this is the *only* $\mathbb{F}[X]$-module structure on V extending the $\mathbb{F}$-vector space structure on V and having $X.v = Tv$. We call the $\mathbb{F}[X]$-module V the **endomorphism module associated to** T

Changing the ring of scalars

Theorem 6.14 (Restriction of scalars)**.** *Suppose that M is an S-module and $\phi : R \to S$ is a ring homomorphism. Then the map $R \times M \to M; (r, x) \mapsto \phi(r).x$ gives M the structure of an R-module.*

Proof. Currying the S-module structure on M gives us a ring homomorphism $\psi : S \to \mathrm{Hom}(M, M)$, and so $\psi \circ \phi$ is a ring homomorphism $R \to \mathrm{Hom}(M, M)$ which uncurries to give an R-module structure on M. In this structure $r.x = \psi \circ \phi(r)(x) = \psi(\phi(r))(x) = \phi(r).x$ as claimed. □

Remark 6.15. The R-module structure above is the result of the **restriction of the scalars of (the S-module) M to R (along ϕ)**.

Remark 6.16. If R is a subring of S then restricting the scalars to R along the inclusion map $R \to S$ corresponds to restricting the scalar multiplication map on $S \times M \to M$ to $R \times M \to M$, but in general ϕ need not be an inclusion map. Indeed, it may not even be injective.

Example 6.17. Suppose that $\phi : \mathbb{F} \to R$ is a ring homomorphism from a field. The additive group of R has an $\mathbb{F}$-module structure induced by ϕ (Proposition 1.39). This $\mathbb{F}$-module structure is the same as that resulting from restricting the scalars of the R-module R to $\mathbb{F}$ along ϕ.

Example 6.18. Suppose V is the $\mathbb{F}[X]$-module described in Example 6.13. The inclusion map $j : \mathbb{F} \to \mathbb{F}[X]$ is a ring homomorphism and the restriction of scalars on the $\mathbb{F}[X]$-module V to $\mathbb{F}$ along j yields the original $\mathbb{F}$-vector space V.

Fields can arise both as subrings of a ring and as quotients of a ring. As a quotient a field can also be used to endow a module with the structure of a vector space, but in this case we have to take care to make sure the scalar multiplication is compatible.

Proposition 6.19. *Suppose that M is an R-module, and I is an ideal in R such that $r.x = 0_M$ for all $r \in I$ and $x \in M$. Then the additive group of M can be given the structure of an R/I-module in such a way that the restriction of scalars to R along the quotient map $q : R \to R/I$ returns the original R-module structure.*

Proof. The scalar multiplication on M curries to a ring homomorphism $\phi : R \to \mathrm{Hom}(M, M)$. By the First Isomorphism Theorem for rings (Theorem 3.31) ϕ induces a well-defined ring homomorphism $\widetilde{\phi} : R/\ker\phi \to \mathrm{Hom}(M, M)$. The zero of the ring $\mathrm{Hom}(M, M)$ is the map $M \to M; x \mapsto 0_M$ and so $\ker\phi = \{r \in R : r.x = 0_M \text{ for all } x \in M\} \supset I$. Hence by Lemma 3.32 the map $\pi : R/I \to R/\ker\phi; x + I \mapsto x + \ker\phi$ is a well-defined ring homomorphism. Uncurrying the ring homomorphism $\widetilde{\phi} \circ \pi$ gives M the structure of an R/I-module with the scalar multiplication $(r + I).x = \widetilde{\phi}(\pi(r + I))(x) = \widetilde{\phi}(r + \ker\phi))(x) = \phi(r)(x) = r.x$. Given this the restriction of scalars of the R/I-module M to R along q recovers the original scalar multiplication on the R-module M as claimed. The result is proved. □

Remark 6.20. Since the map q above is surjective the scalar multiplication of R/I on M is uniquely determined by q: $(r + I).x = r.x$ for all $r \in R$ and $x \in M$ where the first . is the scalar multiplication in the R/I-module M, and the second in the R-module M.

7 Linear maps

As with rings we shall be interested in the structure-preserving maps for modules: An **R-linear map** between two R-modules M and N is a group homomorphism $\phi : M \to N$ with $\phi(r.x) = r.\phi(x)$ for all $x \in M$ and $r \in R$.

Remark 7.1. ⚠ The . on the left is the scalar multiplication on M and the . on the right is the scalar multiplication on N.

Remark 7.2. If $\mathbb{F}$ is a field this has the same meaning as $\mathbb{F}$-linear for vector spaces.

Remark 7.3. Since an R-linear map $\phi : M \to N$ is a group homomorphism, $\phi(0_M) = 0_N$ and $\phi(-x) = -\phi(x)$ for all $x \in M$.

Example 7.4. The identity map $\iota_M : M \to M; x \mapsto x$ and the zero map $z : M \to N; x \mapsto 0_N$ are R-linear. They are seen to be group homomorphisms in Theorem 1.51 and Lemma 1.48 respectivelly, and for linearity we then note that $\iota_M(r.x) = r.x = r.\iota_M(x)$ and $z(r.x) = 0_N = r.0_N = r.z(x)$ (Remark 6.3) for all $r \in R$ and $x \in M$.

Example 7.5. In the R-module R the right multiplication maps are R-linear since multiplication is associative.

Example 7.6. For M an R-module and $r \in R$ we write $r.M$ for the image of the map $\phi : M \to M; x \mapsto r.x$. If R is commutative then ϕ is R-linear since $\phi(x + y) = r.(x + y) = \phi(x) + \phi(y)$ and $\phi(s.x) = r.(s.x) = (rs).x = (sr).x = s.(r.x) = s.\phi(x)$. This example is extremely important and captures many of the ways in which we shall use commutativity.

⚠ This argument fails when R is not commutative.

Lemma 7.7. *Suppose that $\phi, \pi : M \to N$ and $\psi : N \to P$ are R-linear. Then $-\phi$, $\phi + \pi$, and $\psi \circ \phi$ are all R-linear.*

Proof. The fact that all three maps are group homomorphisms is in Lemma 1.48. For linearity it then suffices to note that $(-\phi)(r.x) = -(\phi(r.x)) = -(r.\phi(x)) = r.(-\phi(x)) = r.(-\phi)(x)$ by Remark 6.3; $(\phi+\pi)(r.x) = \phi(r.x)+\pi(r.x) = r.\phi(x)+r.\pi(x) = r.(\phi(x)+\pi(x)) = r.(\phi+\pi)(x)$ for all $r \in R$ and $x \in M$; and $(\psi \circ \phi)(r.x) = \psi(\phi(r.x)) = \psi(r.\phi(x)) = r.\psi(\phi(x)) = r.(\psi \circ \phi)(x)$ for all $r \in R$ and $x \in M$. □

Remark 7.8. Write $\mathrm{End}_R(M)$ for the set of R-linear maps $M \to M$. By the subring test and Example 7.4 and Lemma 7.7, $\mathrm{End}_R(M)$ is a subring of $\mathrm{Hom}(M, M)$.

Remark 7.9. The subgroup test and Lemma 7.7 ensure that the set $L(M, N)$ of R-linear maps $M \to N$ is a commutative group under addition, and as such it has the structure of a $\mathbb{Z}$-module (Example 6.11). ⚠ It does *not*, in general, have the structure of an R-module.

For $\phi \in L(M, N)$ define the function $r.\phi : M \to N$ by $(r.\phi)(x) = r.(\phi(x))$ for all $x \in M$. When R is commutative the map $r.\phi$ is itself R-linear (as a composition of ϕ with a linear

function Example 7.6) and . so-defined gives $L(M,N)$ the structure of an R-module, but if R is not commutative this argument does not work.

Lemma 7.10. *Suppose that $\phi : M \to N$ is an S-linear map, and $\psi : R \to S$ is a ring homomorphism. Then ϕ is an R-linear map between the R-module M with scalars restricted to R along ψ, and the R-module N with scalars restricted to R along ψ.*

Proof. ϕ is a homomorphism of the groups so it suffices to note that $\phi(r.x) = \phi(\psi(r).x) = \psi(r).\phi(x) = r.\phi(x)$. ⚠ All four scalar multiplications here are different: the first is the scalar multiplication of R on M resulting from restricting the scalar multiplication of the S-module M to R along ψ; the second is the scalar multiplication of S on M; the third of S on N; and the fourth is the scalar multiplication of R on N resulting from restricting the scalar multiplication of the S-module N to R along ψ. □

Isomorphisms of modules

We say that $\phi : M \to N$ is an **R-linear isomorphism** if it is an R-linear bijection.

Lemma 7.11. *Suppose that $\phi : M \to N$ is an R-linear isomorphism. Then ϕ^{-1} is R-linear, and hence an R-linear isomorphism.*

Proof. ϕ^{-1} is a group homomorphism since ϕ is a bijective group homomorphism. Hence it is enough to note that $\phi^{-1}(r.x) = \phi^{-1}(r.\phi(\phi^{-1}(x))) = \phi^{-1}(\phi(r.\phi^{-1}(x))) = r.\phi^{-1}(x)$ for all $x \in M$ and $r \in R$ by the R-linearity of ϕ and the fact that ϕ^{-1} is a left and right inverse for ϕ. □

We write $M \cong N$ if there is an R-linear isomorphism $M \to N$.

Proposition 7.12. *$\cong$ is an equivalence relation.*

Proof. The identity map on an R-module is an R-linear isomorphism so $\cong$ is reflexive. $\cong$ is symmetric in view of Lemma 7.11. Finally, $\cong$ is transitive since the composition of bijections is a bijection, and composition of R-linear maps is R-linear – this is Lemma 7.7. □

Example 7.13. Suppose that R is a ring and $z \in U(R)$ then the map $\phi : R \to R; x \mapsto xz$ is an R-linear isomorphism of the R-module R (to itself) - ϕ is a bijection since $R \to R; x \mapsto xz^{-1}$ is an inverse, and ϕ is R-linear as noted in Example 7.5.

Remark 7.14. The R-linear isomorphisms in Example 7.13 are not ring isomorphisms unless $z = 1$. In the other direction, complex conjugation is a ring isomorphism from the ring $\mathbb{C}$ to itself that is not $\mathbb{C}$-linear as a map from the $\mathbb{C}$-module $\mathbb{C}$ to itself.

Submodules

An R-module N is a **submodule** of an R-module M if the map $j : N \to M; x \mapsto x$ is a well-defined R-linear map. We write $N \leqslant M$ and also say that N is **proper** if $M \neq N$.

Remark 7.15. When $R = \mathbb{F}$ is a field so that M is a vector space (see Example 6.4), a submodule of M is exactly a subspace of M.

Lemma 7.16 (Submodule test)**.** *Suppose that M is an R-module and $\varnothing \neq N \subset M$ has $x + y \in N$ for all $x, y \in N$, and $r.x \in N$ whenever $x \in N$ and $r \in R$. Then addition on M and scalar multiplication of R on M restrict to well-defined operations on N giving it the structure of a submodule of M.*

Proof. First, $-1 \in R$ and $(-1).x = -x$ for all $x \in M$ so that by the hypotheses, N is non-empty and $x - y \in N$ whenever $x, y \in N$. It follows that N with addition on M restricted to N, is a subgroup of M by the subgroup test. Since $r.x \in N$ whenever $r \in R$ and $x \in N$, scalar multiplication of R on M restricts to a well-defined function $R \times N \to N$ which *a fortiori* satisfies (M1)–(M4). Finally, the inclusion map is R-linear and the result is proved. □

Remark 7.17. As with rings (Remark 1.33), given a subset satisfying the hypotheses of the above lemma, we make the common abuse of calling it a submodule on the understanding that we are referring to the induced operations.

Example 7.18. Given an R-module M, the zero R-module $\{0\}$ and M itself are submodules of M.

Proposition 7.19. *Suppose that R is a ring. If I is an ideal in the ring R, then I is a submodule of the R-module R. If R is commutative, then if I is a submodule of the R-module R then it is an ideal in the ring R.*

Proof. First, I is an ideal so it is non-empty and closed under addition. Moreover, if $r \in R$ and $x \in I$ then $r.x = rx \in I$, again since I is an ideal and so I is a submodule by the submodule test.

In the other direction, if I is a submodule of R then I is a subgroup of the additive group of the R-module R which is the same as the additive group of the ring R. Moreover, if $r \in R$ and $x \in I$ then $rx = r.x \in I$ since I is a submodule of the R-module R. Finally, since R is commutative $xr = rx \in I$ and so I is an ideal. □

Suppose that M is an R-module and N is a submodule. Then the **annihilator** of N is the set

$$\mathrm{Ann}_R(N) := \{r \in R : r.x = 0_M \text{ for all } x \in N\}.$$

Remark 7.20. With this notation the hypothesis on I in Proposition 6.19 is just $I \subset \mathrm{Ann}_R(M)$.

Proposition 7.21. *Suppose that M is an R-module and N is a submodule of M. Then $\mathrm{Ann}_R(N)$ is an ideal in the ring R.*

Proof. First, $0_R.x = 0_M$ for all $x \in M$ and so $\mathrm{Ann}_R(N)$ is non-empty. Secondly, if $r, s \in \mathrm{Ann}_R(N)$ and $x \in N$ then $(r-s).x = r.x - s.x = 0 - 0 = 0$ so $r - s \in \mathrm{Ann}_R(N)$, and $\mathrm{Ann}_R(N)$ is a subgroup of the additive group of R. Finally, if $s \in \mathrm{Ann}_R(N)$ and $r \in R$ then $(rs).x = r.(s.x) = r.0 = 0$ for all $x \in N$ so $rs \in \mathrm{Ann}_R(N)$, and since N is a module $r.x \in N$ and so $(sr).x = s.(r.x) = 0$ for all $x \in N$ so $sr \in \mathrm{Ann}_R(N)$. The result is proved. □

Quotients and the First Isomorphism Theorem

Theorem 7.22 (Quotient modules)**.** *Suppose that M is an R-module and N is a submodule of M. Then the commutative group M/N may be endowed with the structure of an R-module such that $q : M \to M/N; x \mapsto x + N$ is an R-linear surjection with kernel N.*

Proof. Since N is a commutative subgroup of M we have that M/N is a commutative group and the map q is a surjective homomorphism with kernel N by definition of the quotient group. Define a scalar multiplication of R on M/N by $r.(x+N) := r.x + N$. This is well-defined: if $x + N = y + N$ then $x + n = y + n'$ for some $n, n' \in N$, so $r.x + r.n = r.y + r.n'$, but since N is a submodule $r.n, r.n' \in N$ and hence $r.x + N = r.y + N$ as required.

(M1) follows since $1.(x+N) = (1.x) + N = x + N$ for all $x \in M$ by (M1) for the scalar multiplication on M. (M2) follows since $r.(s.(x+N)) = r.(s.x+N) = (r.(s.x)) + N = (rs).x + N = (rs).(x+N)$ for all $r, s \in R$ and $x \in M$ by (M2) for the scalar multiplication on M. (M3) follows by (M3) for the scalar multiplication on M and the fact that q is a homomorphism so $(r+s).(x+N) = (r+s).x+N = ((r.x)+(s.x))+N = (r.x+N)+(s.x+N) = r.(x+N)+s.(x+N)$ for all $r, s \in R$ and $x \in M$. Finally, (M4) follows by (M4) for the scalar multiplication on M and the fact that q is a homomorphism so $r.((x+N)+(y+N)) = r.((x+y)+N) = r.(x+y)+N = ((r.x)+(r.y))+N = (r.x+N)+(r.y+N)$ for all $r \in R$ and $x, y \in M$.

Finally, it remains to note that q is R-linear by definition and the result is proved. □

Remark 7.23. Since the map q above is a surjective R-linear map the scalar multiplication on M/N is determined by q: $r.(x+N) = r.x + N$ for all $x \in M$ and $r \in R$, where the first . is scalar multiplication in M/N, and the second in M.

By *the* R-module M/N we mean the module structure of this theorem.

Remark 7.24. If $R = \mathbb{F}$ is a field then Theorem 7.22 is exactly the construction of quotient spaces for vector spaces.

Given an R-linear map $\phi : M \to N$, its **kernel** is its kernel as a homomorphism of groups.

Theorem 7.25 (First Isomorphism Theorem for modules)**.** *Suppose that $\phi : M \to N$ is R-linear. Then $\ker\phi$ is a submodule of M; $\operatorname{Im}\phi$ is a submodule of N; and the map*

$$\widetilde{\phi} : M/\ker\phi \to N; x + \ker\phi \mapsto \phi(x)$$

is an injective R-linear map with image $\operatorname{Im}\phi$. In particular, $\operatorname{Im}\phi \cong M/\ker\phi$.

Proof. Both $\ker\phi$ and $\operatorname{Im}\phi$ are subgroups of the additive groups of M and N respectively by the First Isomorphism Theorem for groups since ϕ is, in particular, a group homomorphism. Therefore by the submodule test $\ker\phi$ and $\operatorname{Im}\phi$ are submodules since if $x \in \ker\phi$ then $0_N = r.0_N = r.\phi(x) = \phi(r.x)$ and so $r.x \in \ker\phi$; and if $x \in \operatorname{Im}\phi$ then there is $y \in M$ such that $x = \phi(y)$ and so $r.x = r.\phi(y) = \phi(r.y) \in \operatorname{Im}\phi$.

By Theorem 7.22 $M/\ker\phi$ is an R-module. $\widetilde{\phi}$ is an injective well-defined group homomorphism by the First Isomorphism Theorem for groups. It remains to check that it is linear which follows since $\widetilde{\phi}(r.(x+\ker\phi)) = \widetilde{\phi}((r.x)+\ker\phi) = \phi(r.x) = r.\phi(x) = r.\widetilde{\phi}(x+\ker\phi)$ for all $r \in R$ and $x \in M$. □

Remark 7.26. ⚠ While kernels of ring homomorphisms need not be subrings, kernels of linear maps *are* submodules.

As with rings (*c.f.* Lemma 3.32) it can be useful to be able to quotient by submodules other than the whole kernel in the First Isomorphism Theorem.

Lemma 7.27. *Suppose that M is an R-module and $N \subset P$ are submodules of R. Then the map $R/N \to R/P; x + N \mapsto x + P$ is a well-defined R-linear map.*

Proof. Call the map π. First, π is well-defined since if $x + N = x' + N$ then $x - x' \in N \subset P$ and so $x + P = x' + P$. $\pi((x+N)+(y+N)) = \pi((x+y)+N) = (x+y)+P = (x+P)+(y+P) = \pi(x+N)+\pi(y+N)$, and $\pi(r.(x+N)) = \pi(r.x+N) = r.x+P = r.(x+P)$. The result is proved. □

Example 7.28. The First Isomorphism Theorem applied to the R-linear map $M \to M; x \mapsto x$ gives the isomorphism $M/\{0\} \cong M$; *c.f.* Example 3.33.

Example 7.29. The First Isomorphism Theorem applied to the R-linear map $M \to \{0\}; x \mapsto 0$ gives the isomorphism $M/M \cong \{0\}$.

It will be useful later to have a couple of more involved applications of the First Isomorphism Theorem:

Lemma 7.30. *Suppose that R is a commutative ring, M is an R-module, and $r \in R$. Then $r.M$ is a submodule of M. Furthermore, if N is R-linearly isomorphic to M then $r.N$ is R-linearly isomorphic to $r.M$.*

Proof. $r.M$ is the image of the map $M \to M; x \mapsto r.x$ which is R-linear since R is commutative (Example 7.6), and hence by the First Isomorphism Theorem $r.M$ is a submodule of M.

Let $\phi : M \to N$ be an R-linear isomorphism. The maps $\pi : M \to r.M; x \mapsto r.x$ and $\pi' : N \to r.N; x \mapsto r.x$ are R-linear surjections. ϕ is R-linear and surjective so $\pi' \circ \phi$ is R-linear and $\operatorname{Im} \pi' \circ \phi = r.N$. $x \in \ker \pi' \circ \phi$ if and only if $0_N = r.\phi(x) = \phi(r.x)$. ϕ is an injection so $0_N = \phi(r.x)$ if and only if $r.x = 0_M$. In particular $\ker \pi' \circ \phi = \ker \pi$. By the First Isomorphism Theorem applied to $\pi' \circ \phi$ and π we have $r.N = \operatorname{Im} \pi' \circ \phi \cong M / \ker \pi' \circ \phi = M / \ker \pi \cong \operatorname{Im} \pi = r.M$ as required. □

Lemma 7.31. *Suppose that R is a commutative ring, M is a submodule of the R-module R, and $x \in R$. Then $I := \{r \in R : xr \in M\}$ is a submodule of the R-module R and $x.(R/M) \cong R/I$.*

Proof. The map $R \to x.(R/M); r \mapsto x.(r + M)$ is the composition of two R-linear surjections and so an R-linear surjection. The kernel of this map is exactly the set of $r \in R$ such that $xr + M = x.(r + M) = 0_{R/M} = M$ which is to say I, and the result follows by the First Isomorphism Theorem. □

Direct sums

We defined the direct product of a finite family of rings in Remark 1.56. Its additive group is the direct product of the additive groups of the rings in the family, and the construction extends to infinite families of rings. The direct sum and direct product of a finite family of commutative groups is the same, but for an infinite family of rings there need not be a ring whose additive group is the direct sum of the additive groups of the rings in the family essentially because this set is too small to contain the multiplicative identity.

For modules one can define the direct product of a family of modules, but in this case the additive group of the module is too large if the family is infinite. Here the right generalisation for our interests is the direct sum, and this is why in this section we speak of the direct sum of a finite family of commutative groups, while for rings we spoke of the direct product, even though they are the same for the families we are considering.

Proposition 7.32. *Suppose that R is a ring and $M_1, \ldots, M_n$ are R-modules. Then the commutative group $M_1 \oplus \cdots \oplus M_n$ equipped with the map . defined by $r.x := (r.x_1, \ldots, r.x_n)$ is an R-module.*

Proof. The proposed scalar product is well-defined (the . in the term $r.x_i$ refers to the scalar product on M_i). (M1) follows from (M1) for each of the M_is since $1.x = (1.x_1, \ldots, 1.x_n) = x$. (M2) follows from (M2) for each of the M_is since we have $r.(s.x) = r.(s.x_1, \ldots, s.x_n) = (r.(s.x_1), \ldots, r.(s.x_n)) = ((rs).x_1, \ldots, (rs).x_n) = (rs).x$. (M3) follows from (M3) for each of the M_is and the fact that addition in the group $M_1 \oplus \cdots \oplus M_n$ is coordinatewise, so $(r+s).x =$

$((r+s).x_1,\dots,(r+s).x_n) = (r.x_1+s.x_1,\dots,r.x_n+s.x_n) = (r.x_1,\dots,r.x_n)+(s.x_1,\dots,s.x_n) = r.x+s.x$. Finally, (M4) follows from (M4) for each of the M_is and the fact that addition in the group $M_1\oplus\cdots\oplus M_n$ is coordinatewise, so $r.(x+y) = r.(x_1+y_1,\dots,x_n+y_n) = (r.(x_1+y_1),\dots,r.(x_n+y_n)) = (r.x_1+r.y_1,\dots,r.x_n+r.y_n) = (r.x_1,\dots,r.x_n)+(r.y_1,\dots,r.y_n) = r.x+r.y$. The result is proved. □

Remark 7.33. The R-module denoted $M_1\oplus\cdots\oplus M_n$ is the R-module above and is called the **direct sum of the** M_i**s**. We write M^n for the direct sum of n copies of M with itself.

Direct sums of modules share a number of formal similarities to addition in rings (or modules):

Remark 7.34 (*c.f.* Remark 1.5). For each set I of size n fix a bijection $\sigma:\{1,\dots,n\}\to I$. For R-modules M_i for each $i\in I$, we put an n-module structure on the set $\bigoplus_{i\in I}M_i := \{(x_i)_{i\in I} : x_i\in M_i\}$ via the bijection $\bigoplus_{i\in I}M_i\to M_{\sigma(1)}\oplus\cdots\oplus M_{\sigma(n)}; x\mapsto(x_{\sigma(1)},\dots,x_{\sigma(n)})$. This is what we mean by the R-module $\bigoplus_{i\in I}M_i$; if I is empty then $\bigoplus_{i\in I}M_i$ is the zero module.

For any other bijection $\tau:\{1,\dots,n\}\to I$, $M_{\sigma(1)}\oplus\cdots\oplus M_{\sigma(n)}$ is R-linearly isomorphic to $M_{\tau(1)}\oplus\cdots\oplus M_{\tau(n)}$ via the well-defined R-linear bijection $(x_{\sigma(1)},\dots,x_{\sigma(n)})\mapsto(x_{\tau(1)},\dots,x_{\tau(n)})$, so the particular choice of σ does not change the isomorphism class of $\bigoplus_{i\in I}M_i$.

Remark 7.35 (*c.f.* Remark 1.6). Suppose that I is a finite set and M_i is an R-module for each $i\in I$. If $J\subset I$ is such that $M_i=\{0_{M_i}\}$ for all $i\in I\setminus J$ then $\bigoplus_{i\in I}M_i\to\bigoplus_{i\in J}M_i; x\mapsto(x_i)_{i\in J}$ is a well-defined R-linear isomorphism.

Remark 7.36 (*c.f.* Remark 1.7). If M_i is an R-module for every $i\in I$ and $\mathcal{P}$ is a partition of I then the map

$$\bigoplus_{i\in I}M_i\to\bigoplus_{P\in\mathcal{P}}\left(\bigoplus_{i\in P}M_i\right); x\mapsto((x_i)_{i\in P})_{P\in\mathcal{P}}$$

is a well-defined R-linear isomorphism.

Remark 7.37. Given R-linear maps $\phi_i : M_i\to N_i$ for each $i\in I$, the map $\phi:\bigoplus_{i\in I}M_i\to\bigoplus_{i\in I}N_i; x\mapsto(\phi_i(x_i))_{i\in I}$ is R-linear; if ϕ_i is an injection for all $i\in I$ then ϕ is an injection; if ϕ_i is a surjection for all $i\in I$ then ϕ is a surjection.

Example 7.38. By the R**-module** R^n we mean direct sum of n copies of the R-module R with the convention that it is the zero module if $n=0$.

Example 7.39. The map $R\to R^n; r\mapsto(r,\dots,r)$ is called the **diagonal map** and is an R-linear injection from the R-module R to the R-module R^n.

The Chinese remainder theorem

Theorem 7.40 (Chinese remainder theorem)**.** *Suppose that R is a ring and $M_1,\dots,M_n$ are submodules of the R-module R with $M_j+\bigcap_{i<j}M_i = R$ for all $1<j\leqslant n$. Then the map*

$$\phi:R\to(R/M_1)\oplus\cdots\oplus(R/M_n); r\mapsto(r+M_1,\dots,r+M_n)$$

is a surjective R-linear map with kernel $\bigcap_{i\leqslant n} M_i$. In particular, $R/\bigcap_{i\leqslant n} M_i \cong (R/M_1)\oplus\cdots\oplus(R/M_n)$.

Proof. For $k \leqslant n$ let $\phi_k : R \to (R/M_1)\oplus\cdots\oplus(R/M_k); r \mapsto (r+M_1,\ldots,r+M_k)$. The quotient maps $q_i : R \to R/M_i$ are R-linear by Theorem 7.22, so the map $R^k \to (R/M_1)\oplus\cdots\oplus(R/M_k)$ is R-linear (Remark 7.37), and ϕ_k is the result of composing this with the diagonal map $R \to R^k$ (which is R-linear from Example 7.39), an hence ϕ_k is R-linear. A short check confirms the kernel is $K_k := \bigcap_{i\leqslant k} M_i$. The First Isomorphism Theorem tells us that the map $\psi_k : R/K_k \to (R/M_1)\oplus\cdots\oplus(R/M_k); r+K_k \mapsto (r+M_1,\ldots,r+M_k)$ is a well-defined R-linear isomorphism.

We show by induction for $k \leqslant n$ then ϕ_k is a surjection; for $k = 1$ this is just a property of the quotient map in Theorem 7.22. Assume it is proved for some $k < n$. By hypothesis $M_{k+1} + K_k = R$ and so there is $m \in M_{k+1}$ and $p \in K_k$ such that $m + p = 1$. Suppose $t_1,\ldots,t_{k+1} \in R$, and since ψ_k is an isomorphism let $s \in R$ be such that $\psi_{k+1}(s+K_k) = (t_1+M_1,\ldots,t_k+M_k)$. Then $\phi_k(sm+t_{k+1}p) = (\psi_k(s+(t_{k+1}-s)p+K_k), t_{k+1}+(s-t_{k+1})m+M_{k+1}) = (\psi_k(s+K_k), t_{k+1}+M_{k+1}) = (t_1+M_1,\ldots,t_{k+1}+M_{k+1})$, and so the composition is surjective. The induction is complete.

The very last part of the theorem follows since ψ_n is an R-linear isomorphism. □

Remark 7.41. The history of this theorem is involved – see [She88] – but the starting point is work of Sun Zi (孫子) from around 400AD in which an application of a method for solving simultaneous congruences is given.

Remark 7.42. Ideals I and J in a ring R with $I + J = R$ are said to be **comaximal**. If the submodules M_i in Theorem 7.40 are in fact ideals (Proposition 7.19 tells us this implies that they are submodules of the R-module R) then the hypothesis $M_j + \bigcap_{i<j} M_i = R$ for $1 < j \leqslant n$, can be replaced by the more symmetric requirement that the M_is be pairwise comaximal. Exercise III.6 asks for a proof of this.

Example 7.43. Suppose that p and q are coprime integers then by Bezout's Theorem $\langle p\rangle + \langle q\rangle = \mathbb{Z}$, and we may apply Theorem 7.40 to find that for any $1 \leqslant a \leqslant p$ and $1 \leqslant b \leqslant q$ there is $n \in \mathbb{Z}$ such that $n \equiv a \pmod{p}$ and $n \equiv b \pmod{q}$.

8 Generation and bases

Bases are a crucial tool in vector spaces, and we shall be interested in understanding which modules support an analogue.

Proposition 8.1. *Suppose that M is an R-module and $x_1,\ldots,x_n \in M$. Then the map*

$$\Psi_x : R^n \to M; r \mapsto r_1.x_1 + \cdots + r_n.x_n$$

is R-linear from the R-module R^n.

Proof. It suffices to note that $\Psi_x(r+s) = (r+s)_1.x_1 + \cdots + (r+s)_n.x_n = (r_1+s_1).x_1 + \cdots + (r_n+s_n).x_n = (r_1.x_1 + s_1.x_1) + \cdots + (r_n.x_n + s_n.x_n) = \Psi_x(r) + \Psi_x(s)$ for $r, s \in R^n$, and $\Psi_x(t.r) = (t.r)_1.x_1 + \cdots + (t.r)_n.x_n = (tr_1).x_1 + \cdots + (tr_n).x_n = t.(r_1.x_1) + \cdots + t.(r_n.x_n) = t.\Psi_x(r)$ for all $t \in R$ and $r \in R^n$. □

Remark 8.2. When $n = 0$, Ψ_x just maps 0 to 0_M.

Remark 8.3. By the First Isomorphism Theorem the image of Ψ_x is a submodule of M. We call it the **submodule generated by** $x_1, \ldots, x_n$, and denote it $\langle x_1, \ldots, x_n \rangle$. In particular we write $\langle\rangle = \{0_M\}$.

Example 8.4. If V is a vector space and $x_1, \ldots, x_n \in V$ then the submodule generated by $x_1, \ldots, x_n$ is the subspace spanned by $x_1, \ldots, x_n$.

Remark 8.5. By Corollary 3.15, if R is commutative then the ideal $\langle x_1, \ldots, x_n \rangle$ in the ring R is the same set as the submodule $\langle x_1, \ldots, x_n \rangle$ in the R-module R.

⚠ In Example 3.14 we saw that there is a ring R and an element x such that the ideal generated by x is not equal to Rx, but on the other hand the submodule of the R-module generated by x *is* Rx.

Proposition 8.6. *Suppose that M is an R-module and $\Lambda \subset M$. Then*

$$\langle \Lambda \rangle := \bigcup \{ \langle x_1, \ldots, x_n \rangle : n \in \mathbb{N}_0, x_1, \ldots, x_n \in \Lambda \}$$

is a submodule of M.

Proof. Certainly $0_M \in \langle\rangle \subset \langle \Lambda \rangle$ and so the right hand side is non-empty. If $x, y \in \langle \Lambda \rangle$ then there are $x_1, \ldots, x_n, y_1, \ldots, y_m \in \Lambda$, and $r_1, \ldots, r_n, s_1, \ldots, s_m \in R$ such that $x = r_1.x_1 + \cdots + r_n.x_n$ and $y = s_1.y_1 + \cdots + s_m.y_m$, but then $x + y = r_1.x_1 + \cdots + r_n.x_n + s_1.y_1 + \cdots + s_m.y_m \in \langle x_1, \ldots, x_n, y_1, \ldots, y_m \rangle \subset \langle \Lambda \rangle$. Moreover, if $r \in R$ then $r.x = r.(r_1.x_1 + \cdots + r_n.x_n) = (rr_1).x_1 + \cdots + (rr_n).x_n \in \langle x_1, \ldots, x_n \rangle \subset \langle \Lambda \rangle$ and the result follows by the submodule test. □

Remark 8.7. We call the submodule of this proposition the **module generated by** Λ.

Remark 8.8. For $x_1, \ldots, x_n$ we have $\langle \{x_1, \ldots, x_n\} \rangle = \langle x_1, \ldots, x_n \rangle$.

We say that $x_1, \ldots, x_n$ **generate** M if $M = \langle x_1, \ldots, x_n \rangle$, and Λ is a **generating set for** M if $M = \langle \Lambda \rangle$. If M has a finite generating set then M is said to be **finitely generated**.

Remark 8.9. $x_1, \ldots, x_n \in M$ generate M if and only if the map Ψ_x in Proposition 8.1 is surjective.

Example 8.10. M is a generating set for the R-module M.

Example 8.11. Let $e_i = (0_R, \ldots, 0_R, 1_R, 0_R, \ldots, 0_R)$ be the element of the R-module R^n with a 1_R in the ith position and 0_R elsewhere. $e_1, \ldots, e_n$ generate the R-module R^n since if $r \in R^n$ then $r = r_1.e_1 + \cdots + r_n.e_n$.

Remark 8.12. The generating set of Example 8.11 is a smallest generating set if $R = \mathbb{F}$ – in other words the $\mathbb{F}$-vector space $\mathbb{F}^n$ cannot be generated by strictly fewer than n vectors. This can be bootstrapped to apply to any commutative ring R, but there are rings for which it fails. (See Remark 10.7.)

Proposition 8.13. *Suppose that $\phi : M \to N$ is a surjective R-linear map and $x_1, \ldots, x_n$ generate M. Then $\phi(x_1), \ldots, \phi(x_n)$ generate N.*

Proof. $\Psi_{(\phi(x_1),\ldots,\phi(x_n))} = \phi \circ \Psi_x$ is a composition of surjective maps by hypothesis and so surjective giving the result. □

Remark 8.14. In particular, if M is finitely generated and N is a submodule of M then M/N is finitely generated since the quotient map is a surjection.

Proposition 8.15. *Suppose that $\phi : R \to S$ is a ring homomorphism, and $x_1, \ldots, x_n$ generate the module arising from restricting the scalars of the S-module M to R along ϕ. Then $x_1, \ldots, x_n$ generate the S-module M.*

Proof. Since $x_1, \ldots, x_n$ are a generating set for the R-module arising from restricting the scalars of the S-module M along ϕ, for any $y \in M$ there are elements $r_1, \ldots, r_n \in R$ such that $y = r_1.x_1 + \cdots + r_n.x_n$. By definition $r_i.x_i = \phi(r_i).x_i$ for all i, and so setting $s_i := \phi(r_i)$, we have $y = s_1.x_1 + \cdots + s_n.x_n$ as required. □

Remark 8.16. It is also true (and straightforward to prove) that if ϕ is *surjective* and $x_1, \ldots, x_n$ generate the S-module M then they generate the R-module arising from restricting the scalars of the S-module M along ϕ. This is a natural analogue of Proposition 8.13, but it is not the result we shall use later.

Proposition 8.17. *Suppose that $\phi : M \to N$ is R-linear with $\ker\phi$ generated by a set of size k and $\operatorname{Im}\phi$ generated by a set of size n. Then M is generated by a set of size $n + k$.*

Proof. Let $x_1, \ldots, x_n$ be a generating set for $\operatorname{Im}\phi$ and $y_1, \ldots, y_k$ be a generating set for $\ker\phi$. Since $x_i \in \operatorname{Im}\phi$ there is some $z_i \in M$ such that $\phi(z_i) = x_i$. We shall show that $z_1, \ldots, z_n, y_1, \ldots, y_k$ is a generating set for M: if $v \in M$ then $\phi(v) \in \operatorname{Im}\phi$ and so there are $r_1, \ldots, r_k \in R$ such that

$$\phi(v) = r_1.x_1 + \cdots + r_n.x_n = r_1.\phi(z_1) + \cdots + r_k.\phi(z_n) = \phi(r_1.z_1 + \cdots + r_n.z_n).$$

Thus $v - r_1.z_1 + \cdots + r_n.z_n \in \ker\phi$ and there are elements $s_1, \ldots, s_k \in \ker\phi$ such that $v - r_1.z_1 + \cdots + r_n.z_n = s_1.y_1 + \cdots + s_k.y_k$. The result is proved. □

When a module is generated by a set of size 1 it is said to be **cyclic**.

Example 8.18. A commutative group M is cyclic if and only if the $\mathbb{Z}$-module M (that is M with the $\mathbb{Z}$-module structure described in Example 6.11) is cyclic.

Example 8.19. Suppose that R is a ring and M is a submodule of the R-module R. Then the R-module R/M is cyclic and generated by $1 + M$: indeed, if $r + M \in R/M$ then $r + M = r.(1 + M)$.

Example 8.20. $2\mathbb{Z}$ and $3\mathbb{Z}$ are ideals in the ring $\mathbb{Z}$, and hence (by Proposition 7.19) are submodule so that the quotient $\mathbb{Z}$-modules $\mathbb{Z}/2\mathbb{Z}$ and $\mathbb{Z}/3\mathbb{Z}$ are cyclic $\mathbb{Z}$-modules (Example 8.19). They are *not* isomorphic since they have different sizes.

Generating sets interact particularly well with linear maps.

Proposition 8.21. *Suppose that M is an R-module, Λ generates M, and $\phi, \psi : M \to N$ are R-linear maps with $\phi(e) = \psi(e)$ for all $e \in \Lambda$. Then $\phi = \psi$.*

Proof. Since Λ generates M, for $x \in M$ there are elements $e_1, \dots, e_n \in \Lambda$ and $r_1, \dots, r_n \in R$ such that $x = r_1.e_1 + \cdots + r_n.e_n$, and so

$$\phi(x) = r_1.\phi(e_1) + \cdots + r_n.\phi(e_n) = r_1.\psi(e_1) + \cdots + r_n.\psi(e_n) = \psi(x).$$

The result is proved. □

We say that the $x_1, \dots, x_n \in M$ are **R-linearly independent** if whenever $r_1, \dots, r_n \in R$ have $r_1.x_1 + \cdots + r_n.x_n = 0_M$ we have $r_1, \dots, r_n = 0_R$; we say that a set Λ is **R-linearly independent** if $x_1, \dots, x_n$ is R-linearly independent for every $n \in \mathbb{N}_0$ and *distinct* elements $x_1, \dots, x_n \in \Lambda$.

Remark 8.22. When $R = \mathbb{F}$ is a field our definition of linear independence agrees.

Example 8.23 (Example 8.11, cont.)**.** $e_1, \dots, e_n$ are R-linearly independent: if $r_1, \dots, r_n \in R$ and $r_1.e_1 + \cdots + r_n.e_n = 0$ then $(r_1, \dots, r_n) = 0_{R^n}$ which means $r_1, \dots, r_n = 0$.

Example 8.24. 1 and i are $\mathbb{R}$-linearly independent in the $\mathbb{R}$-module structure induced on $\mathbb{C}$ by the inclusion $\mathbb{R} \to \mathbb{C}$ (Example 1.43), but not $\mathbb{C}$-linearly independent in the $\mathbb{C}$-module $\mathbb{C}$ since $i.1 + (-1).i = 0$.

Example 8.25. The $\mathbb{Z}$-module $\mathbb{Z}/N\mathbb{Z}$ does not contain any non-empty linearly independent sets since $N.x = 0$ for all $x \in \mathbb{Z}/N\mathbb{Z}$.

Remark 8.26. $x_1, \dots, x_n \in M$ are R-linearly independent if and only if the map Ψ_x from Proposition 8.1 is injective.

Remark 8.27. If $\Lambda \subset M$ both generates M and is R-linear independent then we say that Λ is a **basis** for M. The module M is said to be **free** if it has a basis.

Remark 8.28. Every finitely generated vector space has a finite basis and, in particular, is free. In fact assuming the Axiom of Choice, every vector space has a basis [Lan02, Theorem 5.1, Chapter III]. It turns out that the use of the Axiom of Choice is unavoidable in the sense that if every vector space is assumed to have a basis then (in ZF) the Axiom of Choice follows [Bla84, Theorem 1].

Example 8.29. In view of Examples 8.11 & 8.23, $\{e_1, \dots, e_n\}$ is a basis for the R-module R^n, and so the R-module R^n is free.

⚠ An independent subset of the R-module R^n of size n need not be a basis: $\{2\}$ is an independent subset of the $\mathbb{Z}$-module $\mathbb{Z}$ having size 1 but it is not a basis for $\mathbb{Z}$.

Proposition 8.30. *Suppose that M is a finitely generated free R-module. Then M has a finite basis.*

Proof. Let $x_1, \dots, x_n$ generate M, and let $\mathcal{E}$ be a basis. For each i there is a finite $S_i \subset \mathcal{E}$ such that $x_i \in \langle S_i \rangle$, and hence $x_1, \dots, x_n \in \langle \bigcup_{i=1}^n S_i \rangle$. Since $x_1, \dots, x_n$ generate M and $\mathcal{E}$ is R-linearly independent we have $\mathcal{E} \subset \bigcup_{i=1}^n S_i$, and the latter is a finite union of finite sets and so finite. □

Example 8.31. The $\mathbb{Z}$-module $\mathbb{Z}/2\mathbb{Z}$ from Example 8.25 is not free since it is contains non-zero elements, and so is not generated by the empty set, but any non-empty set is not linearly independent.

Example 8.32. The additive group of $\mathbb{Q}$ with the $\mathbb{Z}$-module structure afforded by Example 6.11 is *not* free (see Exercise III.4).

Example 8.33. Suppose that R is a commutative ring. As in Example 6.12, $R[X]$ has the structure of an R-module and in this structure (1.2) and (1.3) exactly say that $\{1, X, X^2, \dots\}$ is a basis.

Remark 8.34. ⚠ A subspace of an n-dimensional vector space is generated by a set of size at most n. There are example of modules (see Exercise IV.3) with a basis of size 1 with submodules that are not even finitely generated, but for PIDs we can recover the situation in the following proposition.

Proposition 8.35. *Suppose that R is a PID and M is a submodule of the R-module R^n. Then M is generated by a set of size at most n.*

Proof. We proceed by induction on n: for $n = 1$, M is a submodule of the R-module R. R is commutative so by Proposition 7.19, M is an ideal in R. Since R is a PID, M is generated

by some $x \in R$ as an ideal in the ring R. Again since R is commutative, the ideal generated by x in the ring R is the same set as the submodule generated by x in the R-module R. Hence M is generated by a set of size 1, as required.

Now suppose the result is true for all submodules of the R-module R^n, and M is a submodule of of the R-module R^{n+1}. The map $\phi : M \to R^n; (x_1, \dots, x_{n+1}) \mapsto (x_1, \dots, x_n)$ is R-linear, so by the First Isomorphism Theorem $\operatorname{Im}\phi$ is a submodule of the R-module R^n. By the inductive hypothesis $\operatorname{Im}\phi$ is generated by a set of size at most n. The map $\psi : \ker\phi \to R; x \mapsto x_{n+1}$ is an R-linear injection (since for $x \in \ker\phi$, $x_1, \dots, x_n = 0$), and so by the First Isomorphism Theorem $\operatorname{Im}\psi$ is a submodule of the R-module R. By the $n = 1$ case $\operatorname{Im}\psi$ is generated by a set of size 1, and $\ker\psi$ is generated by a set of size 0 (since ψ is injective), so by Proposition 8.17 $\ker\phi$ is generated by a set of size 1. Finally, since $\operatorname{Im}\phi$ is generated by a set of size at most n and $\ker\phi$ is generated by a set of size 1, by Proposition 8.17 tells us that M is generated by a set of size $n + 1$. □

Bases are important because they let us realise linear maps as matrices.

Proposition 8.36. *Suppose that R is a ring. For every $A \in M_{n,m}(R)$ the map $L_A : R^n \to R^m; v \mapsto vA$ is R-linear and conversely if $\Phi : R^n \to R^m$ is R-linear then there is $A \in M_{n,m}(R)$ such that $\Phi = L_A$.*

Proof. First, L_A is R-linear since $L_A(v+w) = (vA)+(wA) = (v+w)A$ and $L_A(r.v) = (r.v)A = (rv)A = r(vA)$ by Proposition 1.60. In the other direction let $e_1, \dots, e_n$ be the basis of the R-module R^n defined in Example 8.11, and define $A_{i,1}, \dots, A_{i,m} \in R$ by $\Phi(e_i) = (A_{i,1}, \dots, A_{i,m})$. Then $L_A(e_i) = \Phi(e_i)$ by design and so $L_A = \Phi$ by Proposition 8.21. □

Remark 8.37. A **finite presentation** of a module M is a matrix $A \in M_{n,m}(R)$ and an R-linear isomorphism $R^m/\operatorname{Im} L_A \to M$; M is said to be **finitely presented** if it has a finite presentation. A is said to be a **presentation matrix** for the module M.

Example 8.38. Any finitely generated free module is finitely presented. Indeed, by Proposition 8.30, such a module M has a finite basis, say $x_1, \dots, x_n$, so the map $\Psi_x : R^n \to M$ is an R-linear isomorphism. Let $A \in M_n(R)$ be the zero matrix so that $\operatorname{Im} L_A = \{0\}$, and hence $R^n/\operatorname{Im} L_A \cong R^n \cong M$ as claimed.

Proposition 8.39. *Suppose that R is a PID and M is a finitely generated R-module. Then M is finitely presented.*

Proof. Since M is finitely generated, by Proposition 8.1 there is $m \in \mathbb{N}_0$ and an R-linear surjection $\Psi : R^m \to M$. By the First Isomorphism Theorem this induces an R-linear isomorphism $R^m/\ker\Psi \to M$ and $\ker\Psi$ is a submodule of R^m. By Proposition 8.35 $\ker\Psi$ is finitely generated, and so by Proposition 8.1 there is $n \in \mathbb{N}_0$ and an R-linear surjection

$\Phi : R^n \to R^m$ such that $\operatorname{Im}\Phi = \ker\Psi$. (We can ensure the codomain of Φ is R^m by composing the map afforded by Proposition 8.1 with the inclusion map $\ker\Psi \to R^m$.) By Proposition 8.36 we have $\Phi = L_A$ for some $A \in M_{n,m}(R)$ and we are done. □

Remark 8.40. ⚠ Exercise IV.3 gives an example of a finitely generated module that is not finitely presented.

9 Elementary operations and the Smith normal form

There are three types of **elementary column (resp. row) operation** that can be applied to matrices in $M_{n,m}(R)$ – transvections, dilations, and interchanges – and these correspond to right (resp. left) multiplication by matrices in $M_m(R)$ and $M_n(R)$ respectively.

Remark 9.1. Write $E_n(i,j)$ for the matrix in $M_n(R)$ with 0_Rs everywhere except for row i and column j where the entry is 1_R. Then $E_n(i,j)E_n(k,l) = E_n(i,l)$ if $j = k$ and $E_n(i,j)E_n(k,l) = 0_{M_n(R)}$ if $j \neq k$.

Transvections

Given $1 \leqslant i,j \leqslant m$ with $i \neq j$ and $\lambda \in R$ put $T_m(i,j;\lambda) = I_m + \lambda E_m(i,j)$. Given $A \in M_{n,m}(R)$, the matrix $AT_m(i,j;\lambda)$ is the matrix A with the ith column times λ added to the jth column. We write this

$$A \xrightarrow{c_j \mapsto c_j + c_i\lambda} AT_m(i,j;\lambda).$$

Similarly the matrix $T_n(i,j;\lambda)A$ is the matrix A with λ times the jth row added to the ith row; we write this

$$A \xrightarrow{r_i \mapsto r_i + \lambda r_j} T_n(i,j;\lambda)A.$$

Dilations

Given $1 \leqslant i \leqslant m$ and $u \in U(R)$ let $D_m(i;u) := I_m + (u-1)E_m(i,i)$ so that $D_m(i;u)$ is the matrix with 1s on the diagonal except for the ith element which is u, and 0s elsewhere. The matrix $AD_m(i;u)$ is the matrix with the ith column replaced by the ith column times u and as above we write

$$A \xrightarrow{c_i \mapsto c_i u} AD_m(i;u) \text{ and } A \xrightarrow{r_i \mapsto u r_i} D_n(i;u)A$$

for this and the corresponding row operation.

Interchanges

Given $1 \leqslant i,j \leqslant m$ let $S_m(i,j) = I_m + E_m(i,j) + E_m(j,i) - E_m(i,i) - E_m(j,j)$. The matrix $AS_m(i,j)$ is the matrix A with columns i and j swapped and as above we write

$$A \xrightarrow{c_i \leftrightarrow c_j} AS_m(i,j) \text{ and } A \xrightarrow{r_i \leftrightarrow r_j} S_n(i,j)A$$

for this and the corresponding row operation.

Remark 9.2. We write $\mathrm{GL}_n(R)$ for the group $U(M_n(R))$.

Lemma 9.3. *Suppose that R is a ring. The matrices $T_n(i,j;\lambda)$, $D_n(i;u)$ and $S_n(i,j)$ for $1 \leqslant i,j \leqslant n$, $\lambda \in R$ and $u \in U(R)$ together form an inverse-closed subset of $\mathrm{GL}_n(R)$.*

Proof. These are just calculations: $T_n(i,j;\lambda)T_n(i,j;-\lambda) = I_n = T_n(i,j;-\lambda)T_n(i,j;\lambda)$, so the inverse of a transvection is a transvection. $D_n(i;u)D_n(i;u^{-1}) = I_n = D_n(i;u^{-1})D_n(i;u)$, so the inverse of a dilate is a dilate. Finally, $S_n(i,j)^2 = I_n$ so interchange is self-inverse. □

Remark 9.4. We write $\mathrm{GE}_n(R)$ for the subgroup of $\mathrm{GL}_n(R)$ generated by the matrices in this lemma.

Remark 9.5. In general $\mathrm{GL}_2(R) \neq \mathrm{GE}_2(R)$, though this can be hard to show. An example, taken from [Coh66, p23], is the ring $R := \mathbb{Z}[\theta]$ where $\theta^2 - \theta + 5 = 0$. Here the matrix

$$A := \begin{pmatrix} 3-\theta & 2+\theta \\ -3-2\theta & 5-2\theta \end{pmatrix}$$

is in $\mathrm{GL}_2(R)$ but not in $\mathrm{GE}_2(R)$. This R is a PID as noted in Remark 4.41, where we also stated it is not a Euclidean domain. On the other hand it can be shown that every Euclidean domain has $\mathrm{GL}_2(R) = \mathrm{GE}_2(R)$ so to say that $\mathrm{GL}_2(R) \neq \mathrm{GE}_2(R)$ is a stronger statement. It is an open problem [SZ14, (3), §7] whether every PID with $\mathrm{GE}_2(R) = \mathrm{GL}_2(R)$ is a Euclidean domain.

Remark 9.6. We say that $A, B \in M_{n,m}(R)$ are **equivalent by elementary operations** and write $A \sim_{\mathcal{E}} B$ if there is a sequence $A =: A_0 \to A_1 \to \cdots \to A_{k-1} \to A_k := B$ such that A_{i+1} is the result of an elementary row or column operation applied to A_i for all $0 \leqslant i < k$.

In view of Proposition 1.60 and the definition of $\mathrm{GE}_n(R)$ we have $A \sim_{\mathcal{E}} B$ if and only if there is $P, Q \in \mathrm{GE}_n(R)$ such that $A = PBQ$.

Remark 9.7. We say that $A, B \in M_{n,m}(R)$ are **equivalent** and write $A \sim B$ if there are matrices $S \in \mathrm{GL}_n(R)$ and $T \in \mathrm{GL}_m(R)$ such that $A = SBT$.

Remark 9.8. ⚠ Matrix similarity is a stronger relation than equivalence: First, it required $n = m$; and secondly it requires $Q = P^{-1}$. In other words two matrices A and B are similar if $A, B \in M_n(R)$ and there is $P \in \mathrm{GL}_n(R)$ such that $A = P^{-1}BP$. In particular any square matrix is equivalent to a diagonal matrix (this is Theorem 9.13), but not every square matrix is similar to a diagonal matrix.

Proposition 9.9. *Suppose that R is a ring. Equivalence of matrices is an equivalence relation, and so is equivalence by elementary operations, and the latter is a refinement of the former meaning $A \sim_{\mathcal{E}} B$ implies $A \sim B$.*

Proof. Suppose that $G \leqslant \mathrm{GL}_n(R)$ and $H \leqslant \mathrm{GL}_m(R)$ then the group $G \times H$ acts on $M_{n,m}(R)$ via $(P,Q) * A := PAQ^{-1}$. This is an action because the identity of $G \times H$ is (I_n, I_m) and $I_n A I_m^{-1} = A$, and $((P,Q)(S,T)) * A = (PS, QT) * A = (PS)A(QT)^{-1} = (PS)A(T^{-1}Q^{-1}) = P(SAT^{-1})Q^{-1} = (P,Q) * ((S,T) * A)$ by Proposition 1.60 and the fact, since H is a group, that $(QT)^{-1} = T^{-1}Q^{-1}$. The orbits of an action form a partition, and A and B are the related in the induced equivalence relation if and only if there is $P \in G$ and $Q \in H$ with $A = PBQ^{-1}$. Again, since H is a group and so $(Q^{-1})^{-1} = Q$ this is equivalent to there being $P \in G$ and $Q \in H$ with $A = PBQ$. Taking $G = \mathrm{GL}_n(R)$ and $H = \mathrm{GL}_m(R)$ gives that $\sim$ is an equivalence relation; taking $G = \mathrm{GE}_n(R)$ and $H = \mathrm{GE}_m(R)$ gives that $\sim_{\mathcal{E}}$ is an equivalence relation. The result is proved. □

Equivalence of matrices is useful because of the important result:

Proposition 9.10. *Suppose that R is a ring, $A, B \in M_{n,m}(R)$, and $P \in \mathrm{GL}_n(R)$, $Q \in \mathrm{GL}_m(R)$ are such that $A = PBQ$. Then the map $R^m/\operatorname{Im} L_A \to R^m/\operatorname{Im} L_B; x + \operatorname{Im} L_A \mapsto xQ^{-1} + \operatorname{Im} L_B$ is a well-defined R-linear isomorphism.*

Proof. The map $\phi : R^m \to R^m/\operatorname{Im} L_B; x \mapsto xQ^{-1} + \operatorname{Im} L_B$ is the composition of the linear map $L_{Q^{-1}}$, which is surjective since $L_{Q^{-1}}(xQ) = (xQ)Q^{-1} = x(QQ^{-1}) = xI = x$, and the quotient map $R^m \to R^m/\operatorname{Im} L_B$ which is an R-linear surjection. It follows that ϕ is an R-linear surjection. Furthermore, $\ker\phi = \operatorname{Im} L_A$: if $x \in \ker\phi$ then $xQ^{-1} \in \operatorname{Im} L_B$, which is to say for which there is $v \in R^m$ such that $xQ^{-1} = vB$, whence $x = (vP^{-1})A$ and so $x \in \operatorname{Im} L_A$. Conversely, if $x \in \operatorname{Im} L_A$ then there is $w \in R^m$ such that $x = wA$ hence $xQ^{-1} = wP^{-1}B \in \operatorname{Im} L_B$. The result then follows by the First Isomorphism Theorem. □

Remark 9.11. ⚠ $R^m/\operatorname{Im} L_A \cong R^m/\operatorname{Im} L_B$ does *not* imply that A and B are equivalent. See Exercise IV.4.

Remark 9.12. We say that $A \in M_{n,m}(R)$ is **diagonal** if $A_{i,j} = 0$ for all $i \neq j$. ⚠ The matrix A need not be square to be diagonal.

Theorem 9.13. *Suppose that R is a Euclidean domain. Then every $A \in M_{n,m}(R)$ is equivalent by elementary operations to a diagonal matrix.*

Proof. Let $\mathcal{A}_k$ be those matrices $B \sim_{\mathcal{E}} A$ with the additional property that whenever $i < k$ and $j \neq i$, or $j < k$ and $i \neq j$, we have $B_{i,j} = 0$. We shall show by induction that $\mathcal{A}_k$ is non-empty for $k \leqslant \min\{m,n\}$; $\mathcal{A}_1$ contains A and so is certainly non-empty.

Let f be a Euclidean function for R, and suppose that $\mathcal{A}_k \neq \varnothing$ and $k < \min\{m,n\}$. Let $B \in \mathcal{A}_k$ be a matrix with $f(B_{k,k})$ minimal. First we show that $B_{k,k} \mid B_{k,i}$ for all $i > k$: if not, there is some $i > k$ with $B_{k,i} = qB_{k,k} + r$ with $f(r) < f(B_{k,k})$ and we apply the elementary

operations

$$B = \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,k} & \cdots & B_{k,i} & \cdots & B_{k,m} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & B_{n,k} & \cdots & B_{n,i} & \cdots & B_{n,m} \end{pmatrix}$$

$$\xrightarrow{c_i \mapsto c_i - c_k q} \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,k} & \cdots & B_{k,i} - B_{k,k}q & \cdots & B_{k,m} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & B_{n,k} & \cdots & B_{n,i} - B_{n,k}q & \cdots & B_{n,m} \end{pmatrix}$$

$$\xrightarrow{c_k \leftrightarrow c_i} \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,i} - B_{k,k}q & \cdots & B_{k,k} & \cdots & B_{k,m} \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & B_{n,i} - B_{n,k}q & \cdots & B_{n,k} & \cdots & B_{n,m} \end{pmatrix} =: B'.$$

Then $B' \in \mathcal{A}_k$ has $B'_{k,k} = B_{k,i} - qB_{k,k} = r$, but $f(B'_{k,k}) = f(r) < f(B_{k,k})$ which contradicts the minimality in our choice of B. Similarly, but with row operations in place of column operations, $B_{k,k} \mid B_{i,k}$ for all $i > k$.

For $k < i \leqslant m$ let q_i be such that $B_{k,i} = B_{k,k}q_i$. Apply elementary column operations

$$B \xrightarrow{c_{k+1} \mapsto c_{k+1} - c_k q_{k+1}} \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,k} & 0 & B_{k,k+2} & \cdots & B_{k,m} \\ 0 & \cdots & 0 & B_{k+1,k} & B_{k+1,k+1} - B_{k+1,k}q_{k+1} & B_{k+1,k+2} & \cdots & B_{k+1,m} \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & B_{n,k} & B_{n,k+1} - B_{n,k}q_{k+1} & B_{n,k+2} & \cdots & B_{n,m} \end{pmatrix}$$

$$\vdots$$

$$\xrightarrow{c_m \mapsto c_m - c_k q_m} \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,k} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k+1,k} & B_{k+1,k+1} - B_{k+1,k}q_{k+1} & \cdots & B_{k+1,m} - B_{k+1,k}q_m \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & B_{n,k} & B_{n,k+1} - B_{n,k}q_{k+1} & \cdots & B_{n,m} - B_{n,k}q_m \end{pmatrix} =: B'.$$

For $k < i \leqslant n$ let p_i be such that $B_{i,k} = p_i B_{k,k}$. Apply elementary row operations

$$B' \xrightarrow{r_{k+1} \mapsto r_{k+1} - p_{k+1} r_k} \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,k} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & B'_{k+1,k+1} & \cdots & B'_{k+1,m} \\ 0 & \cdots & 0 & B_{k+1,k} & B'_{k+2,k+1} & \cdots & B'_{k+2,m} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & B_{n,k} & B'_{n,k+1} & \cdots & B'_{n,m} \end{pmatrix}$$

$$\xrightarrow{r_n \mapsto r_n - p_n r_k} \begin{pmatrix} B_{1,1} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & B_{k,k} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & B'_{k+1,k+1} & \cdots & B'_{k+1,m} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & B'_{n,k+1} & \cdots & B'_{n,m} \end{pmatrix} =: B''.$$

Then $B'' \sim_{\mathcal{E}} B' \sim_{\mathcal{E}} B \sim_{\mathcal{E}} A$ and $B'' \in \mathcal{A}_{k+1}$. The inductive step is complete. It follows that $\mathcal{A}_{\min\{m,n\}} \neq \varnothing$; any B in this set is diagonal and equivalent to A. □

Remark 9.14. It is not too hard to adapt the above argument to show that any matrix in a PID is equivalent to a diagonal matrix, however it need not be equivalent by elementary operations. To see this let $R = \mathbb{Z}[\theta]$ and A be as in Remark 9.5. If there were a diagonal matrix Δ with $A \sim_{\mathcal{E}} \Delta$, then there would be $P, Q \in \mathrm{GE}_2(R)$ with $\Delta = PAQ$. The right hand side is invertible so Δ is invertible and therefore $\Delta = D_2(1; u_1)D_2(2; u_2)$ for $u_1, u_2 \in U(R)$ and in particular $\Delta \in \mathrm{GE}_2(R)$. But then $A = P^{-1}\Delta Q^{-1} \in \mathrm{GE}_2(R)$ which gives a contradiction.

Remark 9.15. We say that $A \in M_{n,m}(R)$ is in **Smith normal form** if it is diagonal *and* $A_{i,i} \mid A_{i+1,i+1}$ for all $1 \leqslant i < \min\{n, m\}$.

Proposition 9.16. *Suppose that R is an integral domain in which every finitely generated ideal is principal. Then every diagonal matrix $A \in M_{n,m}(R)$ is equivalent by elementary operations to a matrix in Smith normal form.*

Proof. Let $\mathcal{A}_k$ be the set of diagonal matrices that are elementarily equivalent to A, and such that if the diagonal entries are denoted $a_1, a_2, \ldots, a_{\min\{m,n\}}$, then $a_i \mid a_j$ whenever $1 \leqslant i \leqslant j$ and $i \leqslant k$. Certainly $A \in \mathcal{A}_0$ since the hypotheses on the entries is vacuous then, so there is a maximal $k \in \mathbb{N}^*$ with $k-1 \leqslant \min\{m,n\}$ such that $\mathcal{A}_{k-1}$ is non-empty.

We may assume that $k \leqslant \min\{m,n\}$ since otherwise we are done. By maximality of k for each matrix in $\mathcal{A}_{k-1}$ with diagonal entries $a_1, a_2, \ldots, a_{\min\{m,n\}}$ there is a minimal $l \geqslant k$ with $a_k \nmid a_l$; let $B \in \mathcal{A}_{k-1}$ have l maximal with this property.

Since every finitely generated ideal is principal, $\langle a_k, a_l \rangle = \langle a'_k \rangle$ for some a'_k. Let $\alpha_k, \alpha_l \in R$ be such that $a_k\alpha_l + \alpha_k a_l = a'_k$; write $a_k = a'_k q_k$, $a_l = q_l a'_k$, and $a'_l = -q_l a'_k q_k$. Now

$$B = \begin{pmatrix} \ddots & & & & \\ & a_k & & & \\ & & \ddots & & \\ & & & a_l & \\ & & & & \ddots \end{pmatrix} \xrightarrow{c_l \mapsto c_l + c_k\alpha_l} \begin{pmatrix} \ddots & & & & \\ & a_k & & a_k\alpha_l & \\ & & \ddots & & \\ & & & a_l & \\ & & & & \ddots \end{pmatrix}$$

$$\xrightarrow{r_k \mapsto r_k + \alpha_k r_l} \begin{pmatrix} \ddots & & & & \\ & a_k & & a'_k & \\ & & \ddots & & \\ & & & a_l & \\ & & & & \ddots \end{pmatrix} \xrightarrow{c_k \mapsto c_k - c_l q_k} \begin{pmatrix} \ddots & & & & \\ & 0 & & a'_k & \\ & & \ddots & & \\ & a'_l & & a_l & \\ & & & & \ddots \end{pmatrix}$$

$$\xrightarrow{r_l \mapsto r_l - q_l r_k} \begin{pmatrix} \ddots & & & & \\ & 0 & & a'_k & \\ & & \ddots & & \\ & a'_l & & 0 & \\ & & & & \ddots \end{pmatrix} \xrightarrow{c_l \leftrightarrow c_k} \begin{pmatrix} \ddots & & & & \\ & a'_k & & & \\ & & \ddots & & \\ & & & a'_l & \\ & & & & \ddots \end{pmatrix} =: C$$

The matrix C is diagonal and elementarily equivalent to B, and hence to A; write $a'_1, \ldots, a'_{\min\{m,n\}}$ for its diagonal entries so that for $i \notin \{k,l\}$ we have $a'_i = a_i$. a'_k and a'_l are linear combinations of a_k and a_l and so for $i \leqslant k-1$, a'_i divides them both, and hence for $1 \leqslant i \leqslant j$ we have we have $a'_i \mid a'_j$. It follows that $C \in \mathcal{A}_{k-1}$. Finally $a'_k \mid a_k$ and so $a'_k \mid a'_j$ for $k \leqslant j < l$, but also $a'_k \mid a'_l$ contradicting maximality of l. The result is proved. □

Theorem 9.17. *Suppose that R is a Euclidean domain. Then every $A \in M_{n,m}(R)$ is equivalent by elementary operations to a matrix in Smith normal form.*

Proof. Every Euclidean domain is a PID (Proposition 4.39), and in particular every finitely generated ideal in a PID is principal (since they all are), so the hypotheses of Theorem 9.13 and Proposition 9.16 are satisfied and together they give the result. □

Remark 9.18. Following the work of Kaplanksy [Kap49] an integral domain R for which every $A \in M_{n,m}(R)$ is equivalent to a matrix in Smith normal form, is called an **elementary divisor domain**, so in this language Theorem 9.17 shows that every Euclidean domain is an elementary divisor domain.

Every finitely generated ideal in an elementary divisor domain is principal [LLS74, Theorem 3.1], and it is an open problem [Lor12] (going back at least to [Hel43]) to decide which of the integral domains in which every finitely generated ideal is principal are elementary divisor rings.

10 The structure theorem for modules over Euclidean domains

Theorem 10.1 (Structure of finitely generated modules over a Euclidean domain, invariant factor form)**.** *Suppose that R is a Euclidean domain and M is a finitely generated R-module. Then there are elements $a_1 \mid a_2 \mid \cdots \mid a_n$ with $a_i \not\sim 1$ such that*

$$M \cong (R/\langle a_1\rangle) \oplus \cdots \oplus (R/\langle a_n\rangle),$$

where $\langle a_i\rangle$ denotes the submodule of the R-module R generated by a_i.

Proof. Every Euclidean domain is a PID (Proposition 4.39), so by Proposition 8.39 there is $A \in M_{m,n}(R)$ such that $M \cong R^n/\operatorname{Im} L_A$. If $m < n$ then add $n-m$ rows consisting of zeros to the bottom of the matrix A. This does not change the image of L_A, so even with A modified like this we have $M \cong R^n/\operatorname{Im} L_A$.

By Theorem 9.17 there is a diagonal matrix B with entries $a_1 \mid a_2 \mid \cdots \mid a_{\max\{m,n\}} = a_n$ such that $A \sim B$. By Proposition 9.10 we have $M \cong R^n/\operatorname{Im} L_B$. For $1 \leqslant i \leqslant n$ let M_i be the submodule of the R-module R generated by a_i. By Proposition 7.19 M_i is the same set as the ideal generated by a_i. Finally Theorem 7.22 ensures that the maps $R \to R/M_i; r \mapsto r + M_i$ are R-linear surjections and hence (by Remark 7.37) the map

$$\Phi : R^n \to (R/M_1) \oplus \cdots \oplus (R/M_n); r \mapsto (r_1 + M_1, \ldots, r_n + M_n)$$

is an R-linear surjection. Now $\ker\Phi = M_1 \times \cdots \times M_n$ and $x \in R^m$ has $xB = (a_1x_1, \ldots, a_nx_n)$ since B is diagonal and $m \geqslant n$, so $\ker\Phi = \operatorname{Im} L_B$. It follows by the First Isomorphism Theorem (Theorem 7.25) that $(R/M_1) \oplus \cdots \oplus (R/M_n) \cong R^n/\operatorname{Im} L_B \cong R^n/\operatorname{Im} L_A \cong M$. Finally, if $a_i \sim 1$ then R/M_i is the zero module (Example 7.29) and may be removed by Remark 7.37. The result is proved. □

Theorem 10.2 (Structure of finitely generated modules over a Euclidean domain, primary form)**.** *Suppose that R is a Euclidean domain and M is a finitely generated R-module. Then there are some $s,t \in \mathbb{N}_0$, prime elements $p_1,\dots,p_t \in R$, and $e_1,\dots,e_t \in \mathbb{N}^*$, such that*

$$M \cong R^s \oplus (R/\langle p_1^{e_1}\rangle) \oplus \cdots \oplus (R/\langle p_t^{e_t}\rangle).$$

Proof. Apply Theorem 10.1 to get $a_1 \mid \cdots \mid a_k$ such that $M \cong (R/\langle a_1\rangle)\oplus\cdots\oplus(R/\langle a_k\rangle)$. Since $0 \mid a$ if and only if $a = 0$, we make take $l \in \mathbb{N}_0$ maximal such that $a_l \neq 0$, so that by Example 7.28 (and Remark 7.37) $(R/\langle a_{l+1}\rangle)\oplus\cdots\oplus(R/\langle a_k\rangle) \cong R^{k-l}$. Since R is a UFD (by Proposition 4.39 and Theorem 4.30), we may apply Proposition 4.33 to get $s \in \mathbb{N}_0$ depending only on $a := a_1\cdots a_l$ such that if $f_1,\dots,f_r \nmid 1$ and $f_1\cdots f_r | a$ then $r \leqslant s$. Take r maximal such that there are $f_1,\dots,f_r \nmid 1$ with $f_1\cdots f_r | a$ and $M \cong R/\langle f_1\rangle \oplus \cdots \oplus R/\langle f_r\rangle \oplus R^{k-l}$. Such f_is exist since $a_1,\dots,a_l$ certainly work.

Our aim is to show that for each i there is a prime p_i and an $e_i \in \mathbb{N}^*$ such that $f_i \sim p_i^{e_i}$. f_i is not a unit and R is a UFD, so it has a prime factor p_i. By Proposition 4.33 there is a largest $e_i \in \mathbb{N}^*$ such that $p^{e_i} \mid f_i$. Let q_i be such that $f_i = q_i p_i^{e_i}$. If $c \mid p_i^{e_i}$ then $c \sim p_i^e$, and if c also divides q_i then $e = 0$ by maximality of e_i. It follows that $c \sim 1$ and so $p_i^{e_i}$ and q_i have a greatest common factor and it is c. By Proposition 4.10 we have $\langle p_i^{e_i}\rangle + \langle q_i\rangle = R$. Now $\langle p_i^{e_i}\rangle \cap \langle q_i\rangle$ is principal since R is a PID, say generated by l. By definition l is the lowest common multiple of $p_i^{e_i}$ and q_i and by Proposition 4.18 $f_i = p_i^{e_i} q_i \sim cl \sim l$, hence $\langle p_i^{e_i}\rangle \cap \langle q_i\rangle = \langle f_i\rangle$. Since ideals are submodules in a commutative ring (Proposition 7.19) we can apply the Chinese remainder theorem (Theorem 7.40) to get $R/\langle f_i\rangle \cong (R/\langle p_i^{e_i}\rangle) \oplus (R/\langle q_i\rangle)$. By maximality of r it follows that either $p_i^{e_i} \sim 1$ or $q_i \sim 1$; the former cannot happen since p_i is prime, hence we have the latter and the f_is have the claimed form. □

Perhaps more important than the 'canonical forms' provided by Theorems 10.1 & 10.2 is that they enjoy some uniqueness:

Theorem 10.3. *Suppose that R is a commutative ring, M is an R-module, and $I_1 \subset \cdots \subset I_n$ and $J_1 \subset \cdots \subset J_m$ are proper submodules such that $M \cong (R/I_1)\oplus\cdots\oplus(R/I_n)$ and $M \cong (R/J_1)\oplus\cdots\oplus(R/J_m)$. Then $n = m$ and $J_k = I_k$ for all $1 \leqslant k \leqslant n$.*

Remark 10.4. ⚠ The submodules need to be proper: if $I = R$ then $R/I = \{0_{R/I}\}$ and so by Remark 7.35 $(R/I)^n \cong (R/I)^m$ for all $n,m \in \mathbb{N}_0$.

Lemma 10.5. *Suppose that R is a commutative ring, and $I_1 \subset \cdots \subset I_n$ are submodules of the R-module R. Then $(R/I_1)\oplus\cdots\oplus(R/I_n)$ has a minimal generating set, and it has size k where $k \in \mathbb{N}_0$ is the largest k such that I_k is proper (with $k = 0$ if no I_k is proper).*

Proof. By Remark 7.37 and Example 7.29 we may assume that $k = n$. The map $R^n \to (R/I_1)\oplus\cdots\oplus(R/I_n); (x_1,\dots,x_n) \mapsto (x_1 + I_1,\dots,x_n + I_n)$ is an R-linear surjection from the

R-module R^n by Remark 7.37 and Theorem 7.22. Surjective R-linear maps take generating sets to generating sets (Proposition 8.13), and the R-module R^n has a generating set of size n (Example 8.11). This ensures the first part of the lemma.

Since R is commutative the submodule I_n is an ideal by Proposition 7.19, and so by Theorem 3.50 there is a maximal ideal $J \supset I_n$ (this is where we use that I_n is proper) and hence $J \supset I_k$ for all $1 \leqslant k \leqslant n$. The ideal J is a submodule of the R-module R (Proposition 7.19 in the other direction) and so the map

$$(R/I_1) \oplus \cdots \oplus (R/I_n) \to (R/J)^n; (x_1 + I_1, \ldots, x_n + I_n) \mapsto (x_1 + J, \ldots, x_n + J)$$

is a well-defined R-linear surjection by Lemma 7.27 and Remark 7.37. Suppose that the R-module $(R/I_1) \oplus \cdots \oplus (R/I_n)$ has a generating set of size t. Then by Proposition 8.13 then so does the R-module $(R/J)^n$.

If $r \in J$ and $(x_1+J, \ldots, x_n+J) \in (R/J)^n$ then $r.(x_1+J, \ldots, x_n+J) = (rx_1+J, \ldots, rx_n+J) = (0_{R/J}, \ldots, 0_{R/J}) = 0$ and so we may apply Proposition 6.19, $(R/J)^n$ has the structure of an R/J-module in such a way that the R-module structure arising from restricting the scalars to R along the quotient map $q : R \to R/J$ is the original R-module structure. Hence by Proposition 8.15 this R/J-module has a generating set of size t. Since J is maximal, Proposition 3.46 tells us that R/J is a field, and hence the R/J-module $(R/J)^n$ is an R/J-vector space with a generating set of size t; we conclude (Remark 8.12) that $t \geqslant n$ as required. □

Proof of Theorem 10.3. Since R is commutative, $r.M$ is an R-module for every R (that's the first part of Lemma 7.30). For $r \in R$ write $N_j(r) := \{s \in R : rs \in I_j\}$ for $1 \leqslant j \leqslant n$ which are nested since the I_js are nested, and note that

$$\begin{aligned} & r.M \\ & \cong r.((R/I_1) \oplus \cdots \oplus (R/I_n)) && \text{by Lemma 7.30 since } M \cong (R/I_1) \oplus \cdots \oplus (R/I_n) \text{ (and } R \text{ is commutative)} \\ & = r.(R/I_1) \oplus \cdots \oplus r.(R/I_n) && \text{by definition of scalar multiplication on direct sums} \\ & \cong (R/N_1(r)) \oplus \cdots \oplus (R/N_n(r)). && \text{by Lemma 7.31 } N_j(r) \text{ is a submodule and } r.(R/I_j) \cong R/N_j(r); \text{ the isomorphism is then by Remark 7.37} \end{aligned}$$

It follows by Lemma 10.5 that $r.M$ has a minimal generating set of size $j(r)$ where $j(r)$ is maximal such that $N_{j(r)}(r) \neq R$. If $N_j(r) = R$ then $r = r1_R \in I_j$; if $N_j(r) \neq R$ then $r \notin I_j$. By nesting of the $N_j(r)$s it follows that $j(r)$ is the largest integer such that $r \notin I_{j(r)}$, and 0 if $r \in I_1$. This function j determines the sets $I_1, \ldots, I_n$ but is defined independently of them so we could have just as easily have proceeded for the submodules $J_1, \ldots, J_m$, whence $n = m$ and $I_k = J_k$ for all $1 \leqslant k \leqslant n$. The result is proved. □

Theorem 10.6. *Suppose that R is a commutative ring and the R-modules R^n and R^m are R-linearly isomorphic. Then $n = m$.*

Proof. This follows from Lemma 10.5, Example 7.28 and Remark 7.37. □

Remark 10.7. For vector spaces this is sometimes called the Dimension Theorem. A ring R for which $R^n \cong R^m$ implies $n = m$ is said to have the **invariant basis number (IBN)** property.

11 Applications

Rational canonical form

Given a field $\mathbb{F}$ and a monic polynomial $f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$ we define the $n \times n$ matrix

$$C(f) := \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-2} & -a_{n-1} \end{pmatrix}$$

called the **companion matrix of** f.

Theorem 11.1 (Rational canonical form)**.** *Suppose that V is an n-dimensional vector space over $\mathbb{F}$ and $T : V \to V$ is linear and not identically 0. Then there are monic polynomials $f_1 \mid \cdots \mid f_r$ of degree $n_1, \ldots, n_r$ respectively and with f_1 non-constant, and a basis $v_1, \ldots, v_n$ for V such that $\Psi^{-1}_{v_1,\ldots,v_n} \circ T \circ \Psi_{v_1,\ldots,v_n} = L_M$ where*

$$M = \begin{pmatrix} C(f_1) & 0_{n_1 \times n_2} & \cdots & 0_{n_1 \times n_r} \\ 0_{n_2 \times n_1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{n_{r-1} \times n_r} \\ 0_{n_r \times n_1} & \cdots & 0_{n_r \times n_{r-1}} & C(f_r) \end{pmatrix}.$$

Proof. We regard V as an $\mathbb{F}[X]$-module in the way described in Example 6.13. The restriction of scalars of this module to $\mathbb{F}$ yields the original vector space structure on V (Example 6.18). Since V is finite dimensional as an $\mathbb{F}$-space, this has a finite generating set and hence by Proposition 8.15 the $\mathbb{F}[X]$-module V is finitely generated. Since $\mathbb{F}[X]$ is a Euclidean domain we may apply the Structure Theorem (Invariant Factor Form, Theorem 10.1). We get polynomials $f_1 \mid \cdots \mid f_r$ with $f_1 \not\sim 1$ and

$$\phi : V \to (\mathbb{F}[X]/\langle f_1 \rangle) \oplus \cdots \oplus (\mathbb{F}[X]/\langle f_r \rangle)$$

an $\mathbb{F}[X]$-linear bijection. By Lemma 7.10 the map ϕ is an $\mathbb{F}$-linear bijection but V is finite-dimensional and $\mathbb{F}[X]/\langle 0 \rangle$ is infinite dimensional so $f_i \in \mathbb{F}[X]^*$ for all $1 \leqslant i \leqslant r$. Thus we

may put $n_i \coloneqq \deg f_i$ and may suppose that each f_i is monic (since multiplying by a unit does not change the ideal).

For $1 \leqslant i \leqslant r$ we write $M_i \coloneqq \mathbb{F}[X]/\langle f_i \rangle$ and let $(e_{i,j})_{j=1}^{n_i}$ be such that

$$\phi(e_{i,j}) = (0_{M_1}, \dots, 0_{M_{i-1}}, X^{j-1} + \langle f_i \rangle, 0_{M_{i+1}}, \dots, 0_{M_r}).$$

Then $\phi(e_{1,1}), \dots, \phi(e_{1,n_1}), \phi(e_{2,1}), \dots, \phi(e_{r-1,n_{r-1}}), \phi(e_{r,1}), \dots, \phi(e_{r,n_r})$ is a basis for the $\mathbb{F}$-vector space $M_1 \oplus \cdots \oplus M_r$ and since ϕ is an $\mathbb{F}$-linear isomorphism, the sequence of vectors $e_{1,1}, \dots, e_{1,n_1}, e_{2,1}, \dots, e_{r-1,n_{r-1}}, e_{r,1}, \dots, e_{r,n_r}$ (ordered in this way) is a basis for V as a vector space over $\mathbb{F}$.

Write $f_i(X) = X^{n_i} + a_{n_i-1}^{(i)} X^{n_i-1} + \cdots + a_1^{(i)} X + a_0^{(i)}$ for $1 \leqslant i \leqslant r$. Then since ϕ is $\mathbb{F}[X]$-linear we have

$$\begin{aligned}\phi(Te_{i,j}) = \phi(X.e_{i,j}) = X.\phi(e_{i,j}) &= \begin{cases} \phi(e_{i,j+1}) & \text{if } j < n_i \\ -a_0^{(i)}.\phi(e_{i,1}) - \cdots - a_{n_i-1}^{(i)}.\phi(e_{i,n_i}) & \text{if } j = n_i \end{cases} \\ &= \begin{cases} \phi(e_{i,j+1}) & \text{if } j < n_i \\ \phi(-a_0^{(i)}.e_{i,1} - \cdots - a_{n_i-1}^{(i)}.e_{i,n_i}) & \text{if } j = n_i \end{cases}.\end{aligned}$$

Since ϕ is an $\mathbb{F}$-linear bijection we conclude that T has the required form. □

Remark 11.2. The rational canonical form is also sometimes called the **Frobenius normal form**.

Remark 11.3. The **rational canonical form of a matrix** A is the matrix M from the theorem applied to the linear map L_A.

Jordan normal form

For $\lambda \in \mathbb{C}$ and $n \in \mathbb{N}^*$ define the $n \times n$ matrix

$$J(\lambda, n) \coloneqq \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}.$$

These matrices are called **Jordan blocks**.

Theorem 11.4 (Jordan normal form)**.** *Suppose that V is an n-dimensional vector space over $\mathbb{C}$ and $T : V \to V$ is linear. Then there is a basis $v_1, \dots, v_n$ for V such that $\Psi_{v_1,\dots,v_n}^{-1} \circ$*

$T \circ \Psi_{v_1,\dots,v_n} = L_M$ *where*

$$M = \begin{pmatrix} J(\lambda_1, n_1) & 0_{n_1\times n_2} & \cdots & 0_{n_1\times n_t} \\ 0_{n_2\times n_1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0_{n_{t-1}\times n_t} \\ 0_{n_t\times n_1} & \cdots & 0_{n_t\times n_{t-1}} & J(\lambda_t, n_t) \end{pmatrix}.$$

Proof. We regard V as an $\mathbb{C}[X]$-module in the way described in Example 6.13. The restriction of scalars of this module to $\mathbb{C}$ yields the original vector space structure on V (Example 6.18). Since V is finite dimensional as a $\mathbb{C}$-space, this has a finite generating set and hence by Proposition 8.15 the $\mathbb{C}[X]$-module V is finitely generated. Since $\mathbb{C}[X]$ is a Euclidean domain we may apply the Structure Theorem (Primary Form, Theorem 10.2) to V. We get $s, t \in \mathbb{N}_0$, irreducible polynomials $p_1, \dots, p_t \in \mathbb{C}[X]$, and natural numbers $n_1, \dots, n_t \in \mathbb{N}^*$ such that

$$\phi : V \to (\mathbb{C}[X])^s \oplus (\mathbb{C}[X]/\langle p_1^{n_1}\rangle) \oplus \cdots \oplus (\mathbb{C}[X]/\langle p_t^{n_t}\rangle)$$

is a $\mathbb{C}[X]$-linear bijection. By Lemma 7.10 ϕ is a $\mathbb{C}$-linear bijection but V is finite-dimensional and $\mathbb{C}[X]$ is infinite dimensional so $s = 0$. The irreducible polynomials in $\mathbb{C}[X]$ are all degree 1 (see Example 5.8) thus there are $\lambda_1, \dots, \lambda_t \in \mathbb{C}$ such that $\langle p_i^{n_i}\rangle = \langle (X-\lambda_i)^{n_i}\rangle$; write $M_i := \mathbb{C}[X]/\langle (X-\lambda_i)^{n_i}\rangle$. For each $1 \leqslant i \leqslant t$ let $(e_{i,j})_{j=1}^{n_i}$ be such that

$$\phi(e_{i,j}) = (0_{M_1}, \dots, 0_{M_{i-1}}, (X-\lambda_i)^{j-1} + \langle (X-\lambda_i)^{n_i}\rangle, 0_{M_{i+1}}, \dots, 0_{M_t}).$$

Then $\phi(e_{1,1}), \dots, \phi(e_{1,n_1}), \phi(e_{2,1}), \dots, \phi(e_{t-1,n_{t-1}}), \phi(e_{t,1}), \dots, \phi(e_{t,n_t})$ is a basis for the $\mathbb{C}$-vector space $M_1 \oplus \cdots \oplus M_t$ and since ϕ is a $\mathbb{C}$-linear isomorphism, the sequence of vectors $e_{1,1}, \dots, e_{1,n_1}, e_{2,1}, \dots, e_{t-1,n_{t-1}}, e_{t,1}, \dots, e_{t,n_t}$ (ordered in this way) is a basis for V as a vector space over $\mathbb{C}$.

The map ϕ is $\mathbb{C}[X]$-linear so

$$\phi(Te_{i,j}) = \phi(X.e_{i,j}) = X.\phi(e_{i,j}) = \begin{cases} \phi(e_{i,j+1}) + \lambda_i.\phi(e_{i,j}) & \text{if } j < n_i \\ \lambda_i.\phi(e_{i,j}) & \text{if } j = n_i \end{cases}$$
$$= \begin{cases} \phi(e_{i,j+1} + \lambda_i.e_{i,j}) & \text{if } j < n_i \\ \phi(\lambda_i.e_{i,j}) & \text{if } j = n_i \end{cases}.$$

Since ϕ is a $\mathbb{C}$-linear bijection we conclude that T has the required form. □

Remark 11.5. The fact that every polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$ is vital to the Jordan normal form. We used this fact when we appealed to Example 5.8.

Structure of commutative groups

Theorem 11.6 (Structure of finitely generated commutative groups)**.** *Suppose that G is a finitely generated commutative group. Then there are unique (non-zero) natural numbers $1 \neq d_r \mid d_{r-1} \mid \cdots \mid d_1$ and $s \in \mathbb{N}_0$ such that*

$$G \cong \mathbb{Z}^s \oplus \mathbb{Z}_{d_1} \oplus \cdots \oplus \mathbb{Z}_{d_r}.$$

Proof. G is a $\mathbb{Z}$-module (Example 6.11), so we may apply the Invariant Factor Form of Theorem 10.1 to get the desired structure, writing $\mathbb{Z}^s$ for the s copies of $\mathbb{Z}/\langle 0\rangle$ in the given decomposition by Example 7.29 and Remark 7.37. Then uniqueness follows from the fact that $U(\mathbb{Z}) = \{-1, 1\}$ (Example 2.6) since we have restricted the d_is to be naturals and Theorem 10.3. □

Computing the structure of a commutative group

Suppose that G is a commutative group with generators g_1, g_2, g_3, g_4, g_5 subject to the relations

$$2.g_1 + 6.g_2 - 8.g_3 = 0, g_1 + g_2 + g_4 = 0, \text{ and } 5.g_1 + 5.g_4 + 25.g_5 = 0.$$

What we mean here is that these relations generate the kernel of the map $\mathbb{Z}^5 \to G; z \mapsto z_1.g_1 + \cdots + z_5.g_5$. Writing

$$R := \begin{pmatrix} 2 & 6 & -8 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 5 & 0 & 0 & 5 & 25 \end{pmatrix} \in M_{3,5}(\mathbb{Z}),$$

the map

$$\begin{aligned} G &\to \mathbb{Z}^5/\operatorname{Im} L_R \\ z_1.g_1 + \cdots + z_5.g_5 &\mapsto (z_1, \ldots, z_5) + \operatorname{Im} L_R \end{aligned} \tag{11.1}$$

is a well-defined $\mathbb{Z}$-linear isomorphism by the First Isomorphism Theorem. Our aim is to apply Proposition 9.10 to give a clearer way of viewing G. First, we put R into SNF:

$$R \xrightarrow{r_1 \leftrightarrow r_2} \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 2 & 6 & -8 & 0 & 0 \\ 5 & 0 & 0 & 5 & 25 \end{pmatrix} \xrightarrow[c_4 \mapsto c_4 - c_1]{c_2 \mapsto c_2 - c_1} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 4 & -8 & -2 & 0 \\ 5 & -5 & 0 & 0 & 25 \end{pmatrix}$$

$$\xrightarrow[r_3 \mapsto r_3 - 5r_1]{r_2 \mapsto r_2 - 2r_1} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 4 & -8 & -2 & 0 \\ 0 & -5 & 0 & 0 & 25 \end{pmatrix} \xrightarrow{r_2 \mapsto r_2 + r_3} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -8 & -2 & 25 \\ 0 & -5 & 0 & 0 & 25 \end{pmatrix}$$

$$\xrightarrow{r_3 \mapsto r_3 - 5r_2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -8 & -2 & 25 \\ 0 & 0 & 40 & 10 & -100 \end{pmatrix} \xrightarrow[c_5 \mapsto c_5 + 25c_2]{c_3 \mapsto c_3 - 8c_2 \\ c_4 \mapsto c_4 - 2c_2} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 40 & 10 & -100 \end{pmatrix}$$

$$\xrightarrow{c_3 \leftrightarrow c_4} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 10 & 40 & -100 \end{pmatrix} \xrightarrow[c_5 \mapsto c_5 + 10c_3]{c_4 \mapsto c_4 - 4c_3} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \end{pmatrix}.$$

This tells us there are $P \in \mathrm{GL}_3(\mathbb{Z})$ and $Q \in \mathrm{GL}_5(\mathbb{Z})$ such that

$$P \begin{pmatrix} 2 & 6 & -8 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 5 & 0 & 0 & 5 & 25 \end{pmatrix} Q = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 0 \end{pmatrix} =: S.$$

We can compute Q and R by applying the column (resp. row) operations to the identity since application of column (resp. row) operations corresponds to multiplication by certain matrices and matrix multiplication is associative. In particular, we shall need Q which is just the column operations applied to the identity:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[c_4 \mapsto c_4 - c_1]{c_2 \mapsto c_2 - c_1} \begin{pmatrix} 1 & -1 & 0 & 7 & 0 \\ 0 & 1 & 0 & -8 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[c_5 \mapsto c_5 + 25c_2]{c_3 \mapsto c_3 - 8c_2 \\ c_4 \mapsto c_4 - 2c_2}$$

$$\begin{pmatrix} 1 & -1 & 8 & 1 & -25 \\ 0 & 1 & -8 & -2 & 25 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{c_3 \leftrightarrow c_4} \begin{pmatrix} 1 & -1 & 1 & 8 & -25 \\ 0 & 1 & -2 & -8 & 25 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[c_5 \mapsto c_5 + 10c_3]{c_4 \mapsto c_4 - 4c_3} \begin{pmatrix} 1 & -1 & 1 & 4 & -15 \\ 0 & 1 & -2 & 0 & 5 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -4 & 10 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

By Proposition 9.10 the map

$$G \to \mathbb{Z}^5 / \operatorname{Im} L_S$$
$$z_1.g_1 + \cdots + z_5.g_5 \mapsto zQ + \operatorname{Im} L_S$$

is a well-defined $\mathbb{Z}$-linear isomorphism. Since S is diagonal we view $\mathbb{Z}^5/\operatorname{Im} L_S$ as $\mathbb{Z}/\langle 1\rangle \oplus \mathbb{Z}/\langle -1\rangle \oplus \mathbb{Z}/\langle 10\rangle \oplus \mathbb{Z}/\langle 0\rangle \oplus \mathbb{Z}/\langle 0\rangle$. The first two factors are zero modules and so we get a $\mathbb{Z}$-linear isomorphism

$$G \to \mathbb{Z}_{10} \oplus \mathbb{Z}^2$$
$$z_1.g_1 + \cdots + z_5.g_5 \mapsto (z_1 - 2z_2 + z_4 + 10\mathbb{Z}, 4z_1 + z_3 - 4z_4, -15z_1 + 5z_2 + 10z_4 + z_5).$$

References

[AC11] D. D. Anderson and S. Chun. Irreducible elements in commutative rings with zero-divisors. *Houston J. Math.*, 37(3):741–744, 2011. URL `https://www.math.uh.edu/~hjm/restricted/pdf37(3)/04anderson.pdf`.

[Ber14] D. Berlyne. Ideal theory in rings (Translation of "Idealtheorie in Ringbereichen" by Emmy Noether). 2014, arXiv:1401.2577.

[Bla84] A. Blass. Existence of bases implies the axiom of choice. In *Axiomatic set theory (Boulder, Colo., 1983)*, volume 31 of *Contemp. Math.*, pages 31–33. Amer. Math. Soc., Providence, RI, 1984. doi:10.1090/conm/031/763890.

[Cla10] P. L. Clark. Factorization in integral domains. 2010. URL `http://alpha.math.uga.edu/~pete/factorization2010.pdf`.

[CNT19] C. J. Conidis, P. P. Nielsen, and V. Tombs. Transfinitely valued euclidean domains have arbitrary indecomposable order type. *Communications in Algebra*, 47(3):1105–1113, 2019. doi:10.1080/00927872.2018.1501569.

[Coh66] P. M. Cohn. On the structure of the GL_2 of a ring. *Inst. Hautes Études Sci. Publ. Math.*, (30):5–53, 1966. URL `http://www.numdam.org/item?id=PMIHES_1966__30__5_0`.

[Coh00] P. M. Cohn. *Introduction to Ring Theory*. Springer Undergraduate Mathematics Series. Springer-Verlag London, Ltd., London, 2000. doi:10.1007/978-1-4471-0475-9.

[Con] K. Conrad. Remarks about Euclidean domains. URL `https://kconrad.math.uconn.edu/blurbs/ringtheory/euclideanrk.pdf`.

[Dob16] D. E. Dobbs. On the commutative rings with at most two proper subrings. *International Journal of Mathematics and Mathematical Sciences*, 2016:6912360, 2016. doi:10.1155/2016/6912360.

[Ear14] R. Earl. Groups and Group Actions. Lecture notes, Oxford Prelims, 2014. URL https://courses-archive.maths.ox.ac.uk/node/view_material/43836.

[Fuc58] L. Fuchs. *Abelian groups*. Publishing House of the Hungarian Academy of Sciences, Budapest, 1958.

[Gra74] A. Grams. Atomic rings and the ascending chain condition for principal ideals. *Proc. Cambridge Philos. Soc.*, 75:321–329, 1974. doi:10.1017/s0305004100048532.

[Hel43] O. Helmer. The elementary divisor theorem for certain rings without chain condition. *Bull. Amer. Math. Soc.*, 49(4):225–236, 04 1943. URL https://projecteuclid.org:443/euclid.bams/1183505099.

[Hod79] W. Hodges. Krull implies Zorn. *Journal of the London Mathematical Society*, s2-19(2):285–287, 1979. doi:10.1112/jlms/s2-19.2.285.

[Hun80] T. W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1980. doi:10.1007/978-1-4612-6101-8. Reprint of the 1974 original.

[Jod67] M. A. Jodeit, Jr. Uniqueness in the division algorithm. *Amer. Math. Monthly*, 74:835–836, 1967. doi:10.2307/2315810.

[Kap49] I. Kaplansky. Elementary divisors and modules. *Trans. Amer. Math. Soc.*, 66:464–491, 1949. doi:10.2307/1990591.

[Kea98] M. E. Keating. *A First Course in Module Theory*. Imperial College Press, 1998. doi:https://doi.org/10.1142/p082.

[Lac20] M. Lackenby. Graph Theory, Oxford Part A. 2020. URL https://courses-archive.maths.ox.ac.uk/node/view_material/50953.

[Lam07] T. Y. Lam. *Exercises in modules and rings*. Problem Books in Mathematics. Springer, New York, 2007. doi:10.1007/978-0-387-48899-8.

[Lan02] S. Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2002. doi:10.1007/978-1-4613-0041-0.

[LLS74] M. D. Larsen, W. J. Lewis, and T. S. Shores. Elementary divisor rings and finitely presented modules. *Transactions of the American Mathematical Society*, 187:231–248, 1974. doi:10.2307/1997051.

[Lor12] D. Lorenzini. Elementary divisor domains and Bézout domains. *J. Algebra*, 371:609–619, 2012. doi:10.1016/j.jalgebra.2012.08.020.

[Noe21] E. Noether. Idealtheorie in Ringbereichen. *Math. Ann.*, 83(1-2):24–66, 1921. doi:10.1007/BF01464225.

[Poo19] B. Poonen. Why all rings should have a 1. *Math. Mag.*, 92(1):58–62, 2019. doi:10.1080/0025570X.2018.1538714.

[She88] K. Shen. The historical development of the Chinese remainder theorem. *J. Hangzhou Univ. Natur. Sci. Ed.*, 15(3):270–282, 1988.

[Sou20] K. Soundararajan. Bertrand's postulate and the existence of finite fields. 2020, arXiv:2007.01389.

[SZ14] L. Salce and P. Zanardo. Products of elementary and idempotent matrices over integral domains. *Linear Algebra Appl.*, 452:130–152, 2014. doi:10.1016/j.laa.2014.03.042.

[Tol04] J. R. R. Tolkein. *The Fellowship of the Ring. The Lord of the Rings Part I.* HarperCollins e-books, 50th anniversary edition, 2004. URL `https://s3.amazonaws.com/scschoolfiles/112/j-r-r-tolkien-lord-of-the-rings-01-the-fellowship-of-the-ring-retail-pdf.pdf`.

[Wit31] E. Witt. Über die Kommutativität endlicher Schiefkörper. *Abh. Math. Semin. Univ. Hamb.*, 8:413, 1931. doi:10.1007/BF02941019.

[Zor35] M. Zorn. A remark on method in transfinite algebra. *Bull. Amer. Math. Soc.*, 41(10):667–670, 1935. doi:10.1090/S0002-9904-1935-06166-X.